Appendix A:
Secondary Analysis Results

Page intentionally left blank.

The Secondary Analysis presented in appendix A attempts to identify evidence of the possibility of early treatment effects and initial nonequivalence of the treatment and control groups. In addition, it presents additional analyses to examine whether the findings from the Main Analysis[1] would change if stricter criteria were applied to address nonequivalence.

Early Treatment Effects and Initial Nonequivalence

Early treatment effects could occur during the lag that occurred between the implementation of the curricula at the start of preschool and the baseline pre-testing. For 7 of the 12 research teams, the baseline pretesting began more than 2 weeks after the beginning of the curriculum implementation and for three teams the lag was 5 or more weeks. The models used in the Main Analysis cannot identify positive impacts the curricula might have had on student and classroom measures during this lag period. Therefore, they might underestimate the actual effects of the curricula.

Nonequivalence of the treatment and control group at the baseline pretest could be linked to an early treatment effect or to an unfortunate randomization. If the treatment and control groups were equivalent at the start of the year and the curricula did have a positive effect on the treatment group, then the treatment effect might appear in the baseline pretesting if there had been a long enough lag period. In this case, the treatment group might appear significantly differently than the control at the baseline due to the early treatment effect. Nonequivalence could also occur through an unfortunate randomization of the relatively small number of preschools or classrooms, compared to large scale-up studies, randomized for each curriculum evaluation. In this case, the treatment and control groups might not be equivalent from the start, and there would be relatively low power to detect the nonequivalence. If the treatment and control groups were initially different, then statistically significant differences in their mean post-test results might be due to their initial differences rather than to the impact of a curriculum.

To determine whether there was evidence of nonequivalence and early treatment, the results from the repeated measures models were used. The first step was to identify statistically significant differences between the treatment and control groups' baseline pretest means of each measure. If such a difference was found, the measure was extrapolated back to the beginning of the school year (the start of the treatment). The extrapolation procedure was based in the rate of growth in achievement found during the pre-kindergarten year using the time variable included in the repeated measures model (see appendix B for details). Using an assumption of linear growth over the preschool year (i.e., that the growth rate from the start of the year to the fall pretest was the same as the rate from the pretest to the post-test), the start of year values for the measure were estimated for the treatment and control groups based on their rates of achievement growth and the number of days in the lag period. These start-of-year measures were then statistically tested for equivalency.

If there was a significant difference at the baseline pretest but not at the start of the year, there is some evidence of an early treatment effect (i.e., the groups started out similarly at the beginning of the year but the treatment group made greater gains by the pretest). If there were significant differences at both the baseline pretest and the start-of-year, there is some evidence that the groups were nonequivalent to begin with. Table A-1 identifies the measures for each curriculum that show this type of evidence. The second column identifies any measures that were statistically significantly different at the baseline pretest. The third column notes whether those measures were statistically significantly different at the start of school. The fourth column identifies measures for which there is evidence of an early treatment effect (a significant difference at the pretest and no difference at the start of treatment), and the fifth identifies measures for which there is evidence of nonequivalence at baseline (a significant difference at both the start of treatment and the pretest).

[1] The term "Main Analysis" refers to the analyses presented in chapters 1-13. The term "Secondary Analysis" refers to the analyses presented in appendix A.

Out of the 255 measures examined (17 measures with multiple observations allowing them to be extrapolated backwards multiplied by the 15 curricula), 3 have some evidence of an early treatment effect and 11 have some evidence of nonequivalence.

Table A-1. Possible early treatment effects and non-equivalence at baseline

Curricula	Significant differences at baseline	Significant differences at start of school	Possible early treatment effect	Possible non-equivalence at baseline
Bright Beginnings	ECERS-R Arnett-D	Yes Yes		ECERS-R Arnett-D
Creative Curriculum (Vanderbilt)	ECERS-R Arnett-D	Yes No	Arnett-D	ECERS-R
Creative Curriculum (UNC-Charlotte)				
Creative Curriculum with Ladders to Literacy				
Curiosity Corner	SSRS Problem Behaviors Arnett-P	Yes Yes		SSRS Problem Behaviors Arnett-P
DLM Early Childhood Express with Open Court Reading Pre-K	WJ Letter Word Identification TOLD	Yes No	TOLD	WJ Letter Word Identification
Doors to Discovery	TOLD Arnett-P	Yes Yes		TOLD Arnett-P
Early Literacy and Learning Model				
Language-Focused Curriculum				
Let's Begin with the Letter People	Arnett-P	Yes		Arnett-P
Literacy Express	WJ Letter Word Identification	Yes		WJ Letter Word Identification
Pre-K Mathematics with DLM Early Childhood Express Math software	Shape Composition[1]	No	Shape Composition[1]	
Project Approach				
Project Construct				
Ready, Set, Leap!	Shape Composition[1]	Yes		Shape Composition[1]

[1] Building Blocks, Shape Composition task

NOTE: Arnett-D: Arnett Detachment scale
Arnett-P: Arnett Permissiveness scale

Refer to the glossary for abbreviations of the measures.

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

Considering Initial Nonequivalence in the Analysis

To adjust for the possibility that some of the results of the Main Analysis were affected by initial nonequivalence, a Secondary Analysis was conducted. The Secondary Analysis analyzed the data in two ways. First, the same repeated measures models were used as in the Main Analysis but a stricter criterion was applied to their results making use of the comparison of means at baseline and comparison of growth rates during pre-kindergarten. Second, ANCOVA models were estimated for all measures that had more than one observation (these had been estimated using repeated measures models in the Main Analysis) or a similar preschool baseline measure on a different scale. The earlier observation of the measure was used a covariate when estimating the program impact on the measure observed at a later time. This covariate helped control for possible differences in the measure between the treatment and control groups. The ANCOVA analyses act as a sensitivity analyses to determine whether similar results are obtained using an alternative modeling approach. Both sets of models included the same covariates that were used in the models for the Main Analysis.

Table A-2 identifies which models generated results for which measures. Column three identifies which type of repeated measures model was used and for which grades results were generated. Column four does the same for the ANCOVA models. The repeated measures model could be estimated only for those measures with at least two observations.

Table A-2. Secondary analysis: Outcomes, measures, models, and grades analyzed

Outcome	Measure	Repeated measures model	ANCOVA model with Pre-K baseline covariate
Reading	TERA	Spline: Pre-K and K	Pre-K and K
	WJ Letter Word Identification	Spline: Pre-K and K	Pre-K and K
	WJ Spelling	Spline: Pre-K and K	Pre-K and K
Phonological awareness[1]	Pre-CTOPPP	Simple: Pre-K	Pre-K
	CTOPP		K
Language	PPVT	Spline: Pre-K and K	Pre-K and K
	TOLD	Spline: Pre-K and K	Pre-K and K
Mathematics	WJ Applied Problems	Spline: Pre-K and K	Pre-K and K
	CMA-A	Spline: Pre-K and K	Pre-K and K
	Shape Composition[2]	Spline: Pre-K and K	Pre-K and K
Pre-kindergarten behavior[1]	SSRS Social Skills	Simple: Pre-K	Pre-K
	SSRS Problem Behavior	Simple: Pre-K	Pre-K
	PLBS	Simple: Pre-K	Pre-K
Kindergarten behavior[1]	SSRS Social Skills		K
	SSRS Problem Behavior		K
	LBS		K
Classroom quality	ECERS-R	Simple: Pre-K	Pre-K
Teacher-child interaction	Arnett Detachment	Simple: Pre-K	Pre-K
	Arnett Harshness	Simple: Pre-K	Pre-K
	Arnett Permissiveness	Simple: Pre-K	Pre-K
	Arnett Positive Interaction	Simple: Pre-K	Pre-K

[1] Pre-kindergarten and kindergarten measures are not on the same scale.
[2] Building Blocks, Shape Composition task
NOTE: The repeated measures spline model was used to analyze data collected at three time points (fall and spring of pre-kindergarten and spring of kindergarten). The simple repeated measures model was used to analyze data collected at two time points (fall and spring of pre-kindergarten). Refer to the glossary for abbreviations of the measures.
SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Two criteria were then applied to the results. First, a new one (different for the repeated measures and ANCOVA models) was applied to determine which measures had enough evidence to be considered significant. Second, criteria similar to those used in the Main Analysis were applied to determine the findings on the five student-level outcomes and two classroom-level outcomes. These findings were compared to the findings from the Main Analysis to identify the possibility that nonequivalence may have affected the conclusions of this report.

Secondary Analysis Using Repeated Measures Models

In the Main Analysis, the repeated measures models (both spline and simple) provided a comparison of treatment and control means for the preschool post-test (for all the student-level measures and for six of the classroom-level measures) and a similar comparison for the kindergarten post-test (for eight of the student-level measures). The Secondary Analysis used these same results but in addition took advantage of two other results provided by the repeated measures models: the comparison of the baseline means and the comparison of the growth in achievement during pre-kindergarten. The comparison of the baseline means, if there was no statistically significant difference in the treatment and control group means, gave an initial indication that the groups were equivalent at the time of the pretest. The comparison of the growth in achievement during preschool, if there was a statistically significant greater average growth by the treatment group, provided additional assurance that any statistically significant difference in the post-test means of the treatment and control groups did not reflect initial nonequivalence.

Comparing achievement growth rates in treatment and control groups could only be used for the measures from pre-kindergarten. The repeated measures models tested the growth in achievement from fall pre-kindergarten to spring pre-kindergarten and from spring pre-kindergarten to spring kindergarten. The former tested the impact of the curricula on pre-kindergarten growth in achievement. The latter tested any difference in achievement growth from the end of pre-kindergarten to the end of kindergarten but it did not directly test the curricula's impact on this growth. As a result, the comparison of growth was only used in the Secondary Analysis for the pre-kindergarten results.

Using these additional results from the repeated measures model, the Secondary Analysis required three conditions to be met in order to conclude that a curriculum had a significant effect on a measure for pre-kindergarten: (1) no statistically detectable difference at the pre-kindergarten baseline assessment, (2) a statistically significant covariate-adjusted mean difference between groups at the spring pre-kindergarten post-test, and (3) a statistically significant difference in the rate of growth during pre-kindergarten between the treatment and control groups.

For kindergarten, the Secondary Analysis determined that a curriculum had a significant effect on a measure only if the following two conditions were met: (1) no statistically detectable difference in the pre-kindergarten baseline assessment, and (2) a statistically significant covariate-adjusted mean difference between groups at the spring kindergarten post-test. The lack of the growth comparison made the kindergarten analysis less conservative than the preschool analysis.

Secondary Analysis Using ANCOVA Models

In the Main Analysis, ANCOVA models were used with measures observed only one time. In some cases, similar measures on different scales were observed in pre-kindergarten and in kindergarten so that the pre-kindergarten measure could be included as a covariate in the kindergarten analysis of that measure (e.g., the Comprehensive Test of Phonological and Print Processing [CTOPP]). In the other cases no such covariate existed (e.g., the Teacher Behavior Rating Scale [TBRS]) and the analysis could not control for the initial value of the measure. For the Secondary Analysis, ANCOVA models were used with all the measures for which a similar covariate could be included. This included any measures observed two or three times (which were analyzed with repeated measure models in the Main Analysis) and those measures observed only once but had a similar measure observed in pre-kindergarten (the CTOPP and the three kindergarten behavior measures). Those measures with no similar covariate (the TBRS) were not included in the Secondary Analysis.

The ANCOVA models containing the pre-kindergarten baseline assessment covariate estimate expected means for a given measure at a single time point adjusted for the initial value of that measure. By including an initial value for a measure, the ANCOVA adjusted somewhat for any nonequivalence at the start of treatment although it could not adjust for any differential rates of growth in achievement that resulted from initial differences in the groups. For both pre-kindergarten and kindergarten, the Secondary Analysis concluded that a curriculum had a significant effect on a measure if a statistically significant difference was found in the covariate-adjusted post-test mean from the ANCOVA analyses.

Criteria to Determine Findings

Because of the number of statistical tests that were conducted, some results could be considered significant merely by chance. For example, eight statistical tests were conducted for each of the three reading and math results (start of treatment means, pre-kindergarten fall means, pre-kindergarten spring means, kindergarten spring means, rate of growth fall to spring pre-kindergarten, rate of growth spring pre-kindergarten to spring kindergarten, ANCOVA testing pre-kindergarten spring means, and ANCOVA testing kindergarten spring means) for a total of 24 statistical tests per subject. On average, with alpha at the .05-level, 1.2 tests could be statistically significant by chance. Similarly 16 statistical tests were conducted for the two language measures so 0.8 tests could be statistically significant by chance.

Moreover, within each of the outcomes (mathematics, reading, language, phonological awareness, and behavior) the measures were sufficiently intercorrelated (see table A-3) that an effect on one would not be expected to appear, except by chance, without indications of some effect on the others. Because of the number of tests that were conducted within an outcome and because the measures within an outcome were moderately correlated, criteria were used to decide if *the preponderance of evidence* supported a conclusion that the intervention curriculum resulted in a treatment effect on an outcome by spring of the pre-kindergarten year. These criteria were the same as those used in the Main Analysis.

In practice then, two sets of criteria were applied to the model results for the measures to determine the findings. The first determined whether a curriculum had an impact on a measure. The second determined whether a curriculum had an impact on the student or classroom-level outcomes made up of a group of measures. Table A-4 describes the two criteria. Columns 2 and 3 list the criteria used to determine whether a curriculum affected a measure using either the repeated measures model or the ANCOVA model. Column 4 lists the criteria used to determine whether a curriculum affected an outcome: it is the same criteria used in the Main Analysis.

Table A-3. Correlation matrix for student-level measures

Curricula	WJ Applied Problems	CMA-A Mathematics Composite	Shape Composition	TERA	WJ Letter Word Identification	WJ Spelling	Pre-CTOPPP	PPVT	TOLD	SSRS Social Skills	SSRS Problem Behaviors	PLBS
Mathematics												
WJ Applied Problems	1.00	.67	.47	.67	.52	.48	.63	.70	.68	.21	-.11	.23
CMA-A Mathematics Composite	.67	1.00	.59	.61	.45	.48	.52	.52	.50	.09	-.09	.10
Shape Composition[1]	.47	.59	1.00	.38	.24	.41	.33	.32	.35	.06	-.05	.09
Reading												
TERA	.67	.61	.38	1.00	.70	.62	.46	.63	.55	.15	-.08	.16
WJ Letter Word Identification	.52	.45	.24	.70	1.00	.60	.54	.22	.34	.17	-.13	.16
WJ Spelling	.48	.48	.41	.62	.60	1.00	.29	.34	.31	.10	-.06	.13
Phonological awareness												
Pre-CTOPPP	.63	.42	.32	.54	.31	.29	1.00	.63	.60	.21	-.08	.11
Language												
PPVT	.70	.52	.32	.63	.22	.34	.63	1.00	.52	.19	-.06	.20
TOLD	.68	.50	.35	.55	.34	.31	.60	.70	1.00	.20	-.14	.16
Social skills												
SSRS Social Skills	.21	.09	.06	.15	.16	.10	.21	.19	.20	1.00	-.53	.66
SSRS Problem Behaviors	-.11	-.09	-.05	-.07	-.13	-.06	-.08	-.06	-.14	-.53	1.00	-.75
PLBS	.23	.10	.09	.16	.16	.13	.11	.20	.16	.66	-.75	1.00

[1] Building Blocks, Shape Composition task
NOTE: Refer to the glossary for abbreviations of the measures.
SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Table A-4. Criteria used to determine curricula's impact on a measure and on an outcome

	Criterion 1: Determination that a curriculum affects a measure		Criterion 2: Determination that a curriculum affects an outcome
Grade	**Repeated measures model**	**ANCOVA model**	
Pre-K	(a) No statistically significant difference in the preschool pretest means, and (b) a statistically significant covariate-adjusted mean difference at the preschool post-test, and (c) a statistically significant difference in the rate of growth during preschool	A statistically significant difference in the covariate-adjusted preschool post-test means	For reading, math, and behavior, at least two of the three measures found to be positively affected (and none negatively) At least one of the two language measures found to be positively affected (and none negatively) The phonological awareness measure (Pre-CTOPPP) found to be positively affected The classroom quality measure (ECERS-R) found to be positively affected At least two of the four teacher-child interaction measures found to be positively affected (and none negatively)

NOTE: ANCOVA: Analysis of covariance
Pre-CTOPPP: Preschool Comprehensive Test of Phonological and Print Processing, Elision subtest
ECERS-R: Early Childhood Environment Rating Scale-Revised

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

Results of the Secondary Analysis

Appendix A contains a separate discussion of the Secondary Analysis for each curriculum with the following results (from the repeated measures models unless identified as from the ANCOVA models) provided in table form: (a) covariate adjusted-mean differences at the start of curriculum implementation; (b) covariate-adjusted mean differences at the time of the fall preschool baseline assessment; (c) covariate-adjusted mean differences at the time of the spring preschool post-test assessment (from both the repeated measures and ANCOVA models); (d) the fall to spring pre-kindergarten slope difference (rate of growth from fall to spring of pre-kindergarten); (e) the covariate-adjusted mean differences at the time of the kindergarten post-test assessment (from both the repeated measures and ANCOVA models); and (f) the spring pre-kindergarten to spring kindergarten slope difference (rate of growth from spring pre-kindergarten to spring of kindergarten).

These results are presented in effect size units. Cohen's d was used to provide a measure of the magnitude or size of the treatment effect. For effect sizes calculated at the classroom-level (ES_C), the effect size is the difference between the treatment and control classroom means divided by the pooled standard deviations for classrooms. Because the variation in measures taken at the classroom or group-level tends to be smaller than the variation in measures taken at the individual-level, effect sizes at the classroom-level are generally larger than effect sizes at the student-level. Cohen's d was also used to provide a measure of the slope effect sizes (ES_{Slope}). The slope effect size is the difference between the pre-kindergarten or kindergarten slopes for the treatment and control groups divided by the pooled standard deviations for the child or classroom measure of interest. The slope effect size is a measure of the difference in the rate of growth for the treatment and control groups. See appendix B for more details.

Before turning to the results for the individual curricula, tables A-5 and A-6 summarize the findings from the Secondary Analysis concerning the student and classroom-level outcomes and compare them with the findings from the Main Analysis. Table A-5 provides the findings on the student-level outcomes for the Main and Secondary Analyses and table A-6 provides the findings on the two classroom-level outcomes that could be analyzed under the Secondary Analysis (the four instructional outcomes could not be included). The results in the tables for the Secondary Analysis are footnoted with a "1" if from the repeated measures model and a "2" if from the ANCOVA model.

The tables show that the Secondary Analysis did not identify any curricula affecting the outcomes that were not already identified in the Main Analysis. Also, none of the curricula found to affect outcomes in the Main Analysis affected any additional outcomes under the Secondary Analysis.

The Secondary Analysis reduced the number of impacts found to occur, as would be expected from the application of stricter criteria. Table 5 shows that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* affected preschoolers' reading, phonological awareness, and language under the Main Analysis. Under the Secondary Analysis, it affected preschoolers' reading. *Project Approach* was found to have a negative effect on behavior in the Main Analysis and no effect on behavior in the Secondary Analysis. Table 6 shows that *Creative Curriculum* (UNC-Charlotte) had a positive impact on teacher-child interaction in the Main Analysis and no such effect in the Secondary Analysis. *Lets Begin with the Letter People* was found to have an impact on classroom quality in the Main Analysis but not in the Secondary Analysis.

The other findings from the Main Analysis are similarly found in the Secondary Analysis. *Curiosity Corner* had an effect on kindergarten reading in both analyses. The *Early Literacy and Learning Model (ELLM)* had an effect on language in both analyses. *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* had an effect on kindergarten reading, phonological awareness, and language in both. *Pre-K Mathematics with DLM Early Childhood Express Math software* affected preschool mathematics under both analyses. *Creative Curriculum* (UNC-Charlotte) and *Literacy Express* both had positive impacts on classroom quality in both analyses.

Table A-5. Findings on student-level outcomes: Main and secondary analyses

Curricula	Main analysis					Secondary analysis				
	Reading	Phonological awareness	Language	Math	Behavior	Reading	Phonological awareness	Language	Math	Behavior
Bright Beginnings										
Creative Curriculum (Vanderbilt)										
Creative Curriculum (UNC-Charlotte)										
Creative Curriculum with Ladders to Literacy										
Curiosity Corner	Pre-K: 0 K: +					Pre-K: 0 K: +[1]				
DLM Early Childhood Express with Open Court Reading Pre-K	Pre-K: + K: +	Pre-K: + K: +	Pre-K: + K: +			Pre-K: +[2] K: 0	Pre-K: 0 K: +[2]	Pre-K: 0 K: +[1]		
Doors to Discovery										
Early Literacy and Learning Model			Pre-K: 0 K: +					Pre-K: 0 K: +[1]		
Language-Focused Curriculum										
Let's Begin with the Letter People										
Literacy Express										
Pre-K Mathematics with DLM Early Childhood Express Math software				Pre-K: + K: 0					Pre-K: +[1] K: 0	
Project Approach					Pre-K: 0 K: -					
Project Construct										
Ready, Set, Leap!										

[1] Finding from repeated measures analysis.
[2] Finding from ANCOVA analysis.
NOTE: Abbreviations of the findings are:
Pre-K: Pre-kindergarten
K: Kindergarten
+: Finding of a positive impact
-: Finding of a negative impact
Blank cell: Finding of no impact
0: Finding of no impact (when an impact is found for the other grade)
SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Table A-6. Findings on classroom-level outcomes: Main and secondary analyses

Curricula	Main analysis: Classroom quality	Main analysis: Teacher-child interaction	Secondary analysis: Classroom quality	Secondary analysis: Teacher-child interaction
Bright Beginnings				
Creative Curriculum (Vanderbilt)				
Creative Curriculum (UNC-Charlotte)	+	+	+[1]	
Creative Curriculum with Ladders to Literacy				
Curiosity Corner				
DLM Early Childhood Express with Open Court Reading Pre-K				
Doors to Discovery				
Early Literacy and Learning Model				
Language-Focused Curriculum				
Let's Begin with the Letter People	+			
Literacy Express	+		+[1]	
Pre-K Mathematics with DLM Early Childhood Express Math software				
Project Approach				
Project Construct				
Ready, Set, Leap!				

[1] Finding from ANCOVA analysis

NOTE: Abbreviations of the findings are:

+: Finding of a positive impact

Blank cell: Finding of no impact

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

Bright Beginnings:
Vanderbilt University (Tennessee site)

We present analyses for each of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. The student-level effect sizes (ES_S) and slope effect sizes (ES_{Slope}) are presented in table A-7.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 8 days (including Saturdays, Sundays, and holidays).

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock-Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten scores.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the spring kindergarten assessment or the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was no statistically detectable difference between groups in the covariate-adjusted spring kindergarten means.

For the CMA-A Composite Score, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically significant differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were no statistically detectable differences between groups in the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

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

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included

the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each reading assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the TERA, there was no statistically detectable difference between groups on (a) the covariate-adjusted means at the fall pre-kindergarten assessment, (c) the rate of growth from fall to spring, (d) the rate or growth from fall of pre-kindergarten to spring pre-kindergarten, or (e) the rate of growth from spring pre-kindergarten to spring kindergarten. There was a statistically reliable difference in covariate-adjusted means between groups at the spring pre-kindergarten assessment ($ES_S = .39$, $p < .05$). In this instance, we do not have all three conditions necessary to indicate statistical evidence of a treatment effect on the TERA.

On the ANCOVA, there was no statistically detectable difference between groups in the covariate-adjusted spring pre-kindergarten or spring kindergarten means.

On the WJ Letter Word Identification test, there was (a) no statistically detectable difference in the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was no statistically detectable difference between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically detectable differences in covariate-adjusted means at the fall pre-kindergarten, spring pre-kindergarten, or spring kindergarten assessments, and no statistically detectable differences in rates of growth from fall to spring pre-kindergarten and spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

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

Phonological Awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Tests of Phonological and Print Processing (Pre-CTOPPP). For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted on the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Processing (CTOPP) data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

For the ANCOVA on the kindergarten CTOPP data, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses of the phonological awareness measures, we conclude that *Bright Beginnings* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and the Test of Language Development [TOLD] Grammatic Understanding subtest) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the PPVT, there were no statistically detectable differences between groups in the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the TOLD Grammatic Understanding subtest, there were no statistically detectable differences between groups in the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

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

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and the Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, ANCOVA analyses were conducted on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and the Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences between groups in covariate-adjusted means at (a) the fall pre-kindergarten or (b) spring pre-kindergarten assessments, and (c) no statistically detectable difference in the rate of growth between groups from fall to spring pre-kindergarten. On the SSRS Social Skills scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the SSRS Problem Behaviors measure, there were no statistically detectable differences between groups in covariate-adjusted means at (a) the fall pre-kindergarten or (b) spring pre-kindergarten assessments, and (c) no statistically detectable difference in the rate of growth between groups from fall to spring pre-kindergarten. On the SSRS Problem Behaviors scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten.

On the PLBS, there were no statistically detectable differences between groups in covariate-adjusted means at (a) the fall pre-kindergarten or (b) spring pre-kindergarten assessments, and (c) no statistically detectable difference in the rate of growth between groups from fall to spring pre-kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring pre-kindergarten assessment.

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

Based on the analyses of the behavioral measures, we conclude that *Bright Beginnings* did not have a statistically detectable effect on social and learning behaviors relative to the control condition.

Classroom Outcomes

The classroom-level effect sizes (ES_C) and slope effect sizes (ES_{Slope}) are presented in table A-7.

Overall classroom environment

We obtained observations on the Early Childhood Environment Rating Scale-Revised (ECERS-R) in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with the following covariates: teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R there was a statistically detectable difference in the (a) covariate-adjusted means for the fall pre-kindergarten observation ($ES_C = 1.39$, $p < .05$). The statistically reliable difference on the ECERS-R scale score at the fall observation suggests either the nonequivalence of treatment or control groups or early implementation of the curriculum. To examine the possibility of an effect related to early implementation of the curriculum, we extrapolated back to the beginning of the school year and found a statistically reliable difference between groups ($ES_C = 1.52$, $p < .05$). However, there was not a statistically detectable difference between groups on (b) the spring pre-kindergarten observation, or on (c) the rate of change from the fall to spring pre-kindergarten. On the ANCOVA, there was no statistically detectable difference for the spring pre-kindergarten observation.

Based on the analyses for the ECERS-R, we conclude that *Bright Beginnings* did not have a statistically detectable effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with the following covariates: teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site (note: no fall observation score was included). In addition, for each of the teacher-child relationship measures, ANCOVAs were conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the Arnett Detachment scale, there was a statistically detectable difference in the (a) covariate-adjusted means for the fall pre-kindergarten observation ($ES_C = -1.16$, $p < .05$). The statistically reliable difference on the Arnett Detachment scale score at the fall observation suggests either the nonequivalence of treatment or control groups or early implementation of the curriculum. To examine the possibility of an effect related to early implementation of the curriculum, we extrapolated back to the beginning of the school year and found a statistically reliable difference between groups ($ES_C = -1.47$, $p < .05$). However, there was no statistically detectable difference between groups for (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring pre-kindergarten. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.[2]

On the Arnett Harshness scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Permissiveness scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Positive Interactions scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses for the four teacher-child relationship measures, we conclude that *Bright Beginnings* did not have a statistically detectable effect on classroom environment relative to the control condition.

Classroom instruction

Because the classroom instruction measures (TBRS Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in the spring pre-kindergarten observation, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for *Bright Beginnings*

The impact of *Bright Beginnings* on the child- and classroom-level measures is summarized in table A-7.

[2] Even though there was a statistically significant difference between groups on the extrapolated start of treatment means, on the ANCOVA analysis, which covaries out any differences between groups in the fall assessment, we did not obtain a statistically significant difference between groups in the spring pre-kindergarten assessment.

Table A-7. Secondary analysis results for *Bright Beginnings*

Measure	RM analysis start of treatment[1]	RM analysis Fall Pre-K	RM analysis Spring Pre-K	Fall-Spring slope	ANCOVA[2] Spring Pre-K	RM analysis kindergarten	Spring Pre-K-Spring K slope	ANCOVA kindergarten
				Student-level effect sizes				
Mathematics								
WJ Applied Problems	-.08	-.04	.16	.1931	.18	.13	-.0173	.11
CMA-A Mathematics Composite	-.06	-.02	.14	.1595	.10	.07	-.0420	-.01
Shape Composition[3]	.11	.09	-.03	-.1162	-.07	.15	.0958	.08
Reading								
TERA	.02	.09	.39*	.2908	.32	-.07	-.2478	-.19
WJ Letter Word Identification	.24	.26	.35	.0912	.11	.09	-.1426	-.08
WJ Spelling	-.02	.02	.18	.1581	.20	.06	-.0665	-.02
Phonological awareness								
Pre-CTOPPP/CTOPP	.07	.04	-.07	-.1130	-.08	†	†	.01
Language								
PPVT	-.10	-.05	.13	.1815	.12	.07	-.0345	.04
TOLD	-.15	-.10	.09	.1932	.18	.16	.0344	.14
Behavior								
SSRS Social Skills	-.35	-.33	-.27	.0643	-.12	†	†	-.03
SSRS Problem Behavior[4]	.04	.07	.23	.1493	.19	†	†	.24
PLBS/LBS	.05	.04	.04	-.0059	.03	†	†	-.30
				Classroom-level effect sizes				
Global classroom quality								
ECERS-R	1.52*	1.39*	.80	-.5726	1.53	†	†	†
Teacher-child interaction								
Arnett Detachment[5]	-1.47*	-1.16*	.19	1.3204	.15	†	†	†
Arnett Harshness[5]	-.85	-.67	.12	.7694	-.04	†	†	†
Arnett Permissiveness[5]	-.61	-.47	.16	.6148	.10	†	†	†
Arnett Positive Interactions	.96	.86	.41	-.4368	-.14	†	†	†

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

* $p < .05$

[1] The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).

[2] The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.

[3] Building Blocks, Shape Composition task

[4] Higher scores on this scale represent more negative child behaviors.

[5] Lower scores on this scale represent a more positive classroom environment.

NOTE: RM: Repeated Measures
ANCOVA: Analysis of covariance

Refer to the glossary for abbreviations of the measures.

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

Creative Curriculum: Vanderbilt University (Tennessee site)

Creative Curriculum was evaluated by two research teams—Vanderbilt University (Tennessee) and University of North Carolina at Charlotte (North Carolina). Here we present analyses from the Tennessee site, beginning with the analyses of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. The student-level effect sizes (ES_S) and slope effect sizes (ES_{Slope}) are presented in table A-8.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 8 days (including Saturdays, Sundays, and holidays).

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock-Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition task). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups on the covariate-adjusted spring kindergarten means.

For the CMA-A Composite Score, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups on the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were no statistically detectable differences between groups in the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

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

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each reading assessment, an ANCOVA was conducted in which the covariates were *fall assessment score*, child age, gender, disability status as reported by parent, race/ethnicity, and mother's education.

For the TERA, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically detectable differences in covariate-adjusted means at the fall pre-kindergarten, spring pre-kindergarten, or spring kindergarten assessments, and no statistically detectable differences in rates of growth from fall to spring pre-kindergarten and spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

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

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted for the pre-kindergarten Pre-CTOPPP data and the kindergarten CTOPP data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups in the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

On the ANCOVA for the Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

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

Language assessments

Data from the two language measures (Peabody Picture and Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatic Understanding subtest) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included).

In addition, for each language assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) spring kindergarten assessment, but there was a statistically detectable difference on the (d) rate of growth favoring the *Creative Curriculum* group from fall pre-kindergarten to spring pre-kindergarten (difference in slope = 4.25; $ES_{Slope} = .2414$, $p < .01$). We did not obtain all three conditions necessary to indicate statistical evidence of a treatment effect on the PPVT. On the ANCOVA, there was no statistically detectable difference between groups in the covariate-adjusted means for the spring pre-kindergarten assessment.

For kindergarten, on the PPVT, there were no statistically detectable differences between groups on the (e) spring kindergarten assessment or on the (f) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring kindergarten.

On the TOLD Grammatic Understanding subtest, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten.

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

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, ANCOVA analyses were conducted on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills, SSRS Problem Behaviors, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences between groups in the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Social Skills scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten.

On the SSRS Problem Behaviors measure, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Problem Behaviors scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten.

On the PLBS, there were no statistically detectable differences in covariate-adjusted means on the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth between groups from fall to spring pre-kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring pre-kindergarten assessment.

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

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

Classroom Outcomes

The classroom-level effect sizes (ES_C) and slope effect sizes (ES_{Slope}) are presented in table A-8.

Overall classroom environment

We obtained observations on the Early Childhood Environment Rating Scale-Revised (ECERS-R) in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R, there was statistically detectable difference in the (a) covariate-adjusted means for the fall pre-kindergarten observation ($ES_C = 1.94$, $p < .01$). However, there was no statistically detectable difference between groups for the (b) spring pre-kindergarten observation, (c) the spring kindergarten observation, or (d) the rate of change from the fall to spring observation. The statistically reliable difference on the ECERS-R scale score at the fall observation suggests either the nonequivalence of the treatment and control groups or early implementation of the study curriculum. To examine the possibility of an effect related to early implementation of the curriculum, we extrapolated back to the beginning of the school year and found a statistically reliable difference between groups ($ES_C = 2.28$, $p < .001$). However, there was not a statistically detectable difference between groups for (b) the spring pre-kindergarten observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.[3]

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

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, for each of the teacher-child relationship measures, ANCOVAs were conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the Arnett Detachment scale, there were no statistically detectable differences between the groups on the (a) covariate-adjusted means for the fall pre-kindergarten observation, the (b) spring pre-kindergarten

[3] Even though there was a statistically significant difference between groups on the extrapolated start of treatment means, on the ANCOVA analysis, which covaries out any differences between groups at the fall observation, we did not obtain a statistically significant difference between groups on the spring pre-kindergarten observation.

observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.[4]

On the Arnett Harshness scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Permissiveness scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Positive Interactions scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or (c) rate of change from the fall to spring assessment. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten assessment.

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

Classroom instruction

Because data derived from the classroom instruction measures (Teacher Behavior Rating Scale [TBRS] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in the spring pre-kindergarten assessment, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for *Creative Curriculum* (Tennessee site)

The impact of *Creative Curriculum* on the child- and classroom-level measures is summarized in table A-8.

[4] Even though there was a statistically significant difference between groups on the extrapolated start of treatment means, on the ANCOVA analysis, which covaries out any differences between groups at the fall observation, we did not obtain a statistically significant difference between groups on the spring pre-kindergarten observation.

Table A-8. Secondary analysis results for *Creative Curriculum*: Tennessee

Measure	RM analysis start of treatment[1]	RM analysis Fall Pre-K	RM analysis Spring Pre-K	Fall-Spring slope	ANCOVA[2] Spring Pre-K	RM analysis kindergarten	Spring Pre-K-Spring K slope	ANCOVA kindergarten
				Student-level effect sizes				
Mathematics								
WJ Applied Problems	.09	.10	.17	.0671	.07	.17	-.0013	.09
CMA-A Mathematics Composite	.05	.06	.10	.0363	.03	.05	-.0265	.04
Shape Composition[3]	.13	.09	-.12	-.2036	-.13	.00	.0629	-.04
Reading								
TERA	-.08	-.06	.02	.0791	.06	.10	.0430	.03
WJ Letter Word Identification	.32	.29	.16	-.1281	-.11	.38	.1156	.08
WJ Spelling	-.12	-.06	.19	.2402	.20	.25	.0345	.21
Phonological awareness								
Pre-CTOPPP/CTOPP	-.16	-.15	-.10	.0406	-.01	†	†	.06
Language								
PPVT	-.07	-.01	.23	.2414**	.21	.12	-.0624	.08
TOLD	-.07	-.05	.07	.1109	.09	.11	.0234	.14
Behavior								
SSRS Social Skills	-.27	-.22	-.03	.1847	.09	†	†	.35
SSRS Problem Behavior[4]	.01	.02	.07	.0544	.05	†	†	-.05
PLBS/LBS	-.03	.00	.14	.1357	.13	†	†	.08
				Classroom-level effect sizes				
Global classroom quality								
ECERS-R	2.28**	1.94**	.45	-1.4470	1.57	†	†	†
Teacher-child interaction								
Arnett Detachment[5]	-1.13	-.95*	-.16	.7686	-.16	†	†	†
Arnett Harshness[5]	-.36	-.32	-.12	.1945	-.53	†	†	†
Arnett Permissiveness[5]	-.27	-.13	.51	.6173	.60	†	†	†
Arnett Positive Interactions	.95	.74	-.15	-.8677	-.50	†	†	†

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

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

[1] The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).

[2] The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.

[3] Building Blocks, Shape Composition task

[4] Higher scores on this scale represent more negative child behaviors.

[5] Lower scores on this scale represent a more positive classroom environment.

NOTE: RM: Repeated Measures
ANCOVA: Analysis of covariance

Refer to the glossary for abbreviations of the measures.

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

Creative Curriculum: University of North Carolina at Charlotte (North Carolina and Georgia sites)

Creative Curriculum was evaluated by the University of North Carolina research team and by the Vanderbilt University research team. Here we present the results of the North Carolina research team evaluation. The student-level effect sizes (ES_S) and slope effect sizes (ES_{Slope}) are presented in table A-9.

The North Carolina team implemented *Creative Curriculum* at sites in North Carolina and in Georgia. We present the analyses that combine the two implementation sites. We begin with the analyses of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 16 days (including Saturdays, Sundays, and holidays) for North Carolina and 14 days for Georgia.

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock-Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the CMA-A Composite Score, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was a statistically detectable difference between groups in the covariate-adjusted spring pre-kindergarten means (ES_S = .32, p < .05), and no statistically detectable significant differences between groups in the covariate-adjusted spring kindergarten means. In this instance, we do not have all three conditions necessary to indicate statistical evidence of a treatment effect on Shape Composition at spring pre-kindergarten relative to the control condition.

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

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each reading assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the TERA, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

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

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted on the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically significant differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there was no statistically significant difference between groups in the covariate-adjusted spring pre-kindergarten means.

On the ANCOVA for the CTOPP, there was no statistically detectable difference between groups in the covariate-adjusted spring kindergarten means.

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

Language assessments

Data from the two language measures (Peabody Picture and Vocabulary Test [PPVT] and Test of Early Learning Development [TOLD] Grammatic Understanding subtest) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each language assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the TOLD Grammatic Understanding subtest, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

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

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, we conducted an ANCOVA on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Social Skills scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the SSRS Problem Behaviors scale measure, there were no statistically detectable differences in covariate-adjusted means at the (a) fall pre-kindergarten or (b) spring pre-kindergarten assessments, and (c) no statistically detectable difference in the rate of growth between groups from fall to spring pre-kindergarten. On the SSRS Problem Behaviors scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the PLBS, there were no statistically detectable differences in covariate-adjusted means at (a) the fall pre-kindergarten or (b) spring pre-kindergarten assessments, and (c) no statistically detectable difference in the rate of growth between groups from fall to spring pre-kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring pre-kindergarten assessment.

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

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

Classroom Outcomes

The classroom-level effect sizes (ES_C) and slope effect sizes (ES_{Slope}) are presented in table A-9.

Overall classroom environment

We obtained observations on the Early Childhood Environment Rating Scale-revised (ECERS-R) in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. However, there was a statistically significant difference between groups on the (b) spring pre-kindergarten observation ($ES_C = 1.66$, $p < .05$). On the ANCOVA, a statistically significant difference was obtained on the spring pre-kindergarten observation ($ES_C = 1.36$, $p < .01$). *Creative Curriculum* classrooms received higher global classroom quality ratings relative to the control group classrooms.

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

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, for each of the teacher-child relationship measures, ANCOVAs were conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the Arnett Detachment scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. However, there was a statistically significant difference between groups on (b) the spring pre-kindergarten observation ($ES_C = -1.68$, $p < .05$). *Creative Curriculum* teachers were rated as less detached in their interactions with students relative to teachers in the control classrooms. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Harshness scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Permissiveness scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Positive Interactions scale, there was no statistically significant difference between groups in the (a) covariate-adjusted means for the fall pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. There was a statistically significant difference between groups on the (b) spring pre-kindergarten observation ($ES_C = 1.65$, $p < .01$). Teachers in *Creative Curriculum* classrooms were more positive in their interactions with students relative to teachers in the control classrooms as measured by the Arnett Positive Interactions scale. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses of the teacher-child relationship scales, we conclude that *Creative Curriculum* had a positive effect on teacher-child relationships relative to the control condition in pre-kindergarten but no effect in kindergarten.

Classroom instruction

Because the classroom instruction measures (Teacher Behavior Rating Scale [TBRS] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in spring pre-kindergarten, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for *Creative Curriculum* (North Carolina and Georgia sites)

The impact of *Creative Curriculum* on the child- and classroom-level measures is summarized in table A-9.

Table A-9. Secondary analysis results for *Creative Curriculum*: North Carolina and Georgia

Measure	RM analysis start of treatment[1]	RM analysis Fall Pre-K	RM analysis Spring Pre-K	Fall-Spring slope	ANCOVA[2] Spring Pre-K	RM analysis kindergarten	Spring Pre-K-Spring K slope	ANCOVA kindergarten
				Student-level effect sizes				
Mathematics								
WJ Applied Problems	.17	.17	.20	.0273	.16	.09	-.0579	.08
CMA-A Mathematics Composite	.04	.01	-.10	-.1125	-.05	.14	.1273	.16
Shape Composition[3]	-.03	.01	.19	.1703	.32*	-.01	-.1036	.03
Reading								
TERA	.17	.13	-.08	-.1968	-.20	-.04	.0194	-.16
WJ Letter Word Identification	-.22	-.19	-.08	.1131	.05	.00	.0392	.16
WJ Spelling	.06	.01	-.18	-.1868	-.22	-.05	.0688	.02
Phonological awareness								
Pre-CTOPPP/CTOPP	.00	.00	.02	.0144	.05	†	†	.06
Language								
PPVT	.00	.02	.08	.0595	.11	.15	.0377	.12
TOLD	.21	.14	-.16	-.2884	-.17	-.17	-.0095	-.25
Behavior								
SSRS Social Skills	.22	.19	.05	-.1375	.00	†	†	-.12
SSRS Problem Behavior[4]	-.10	-.11	-.16	-.0468	-.13	†	†	.08
PLBS/LBS	.21	.18	.07	-.1109	.02	†	†	-.20
				Classroom-level effect sizes				
Global classroom quality								
ECERS-R	.33	.58	1.66*	1.0578	1.36**	†	†	†
Teacher-child interaction								
Arnett Detachment[5]	-.40	-.64	-1.68*	-1.0160	-1.25	†	†	†
Arnett Harshness[5]	-.66	-.67	-.70	-.0260	-.18	†	†	†
Arnett Permissiveness[5]	.67	.35	-1.01	-1.3300	-.76	†	†	†
Arnett Positive Interactions	.15	.43	1.65**	1.1926	1.40	†	†	†

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

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

[1] The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).

[2] The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.

[3] Building Blocks, Shape Composition task

[4] Higher scores on this scale represent more negative child behaviors.

[5] Lower scores on this scale represent a more positive classroom environment.

NOTE: RM: Repeated Measures
ANCOVA: Analysis of covariance

Refer to the glossary for abbreviations of the measures.

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

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

We present analyses for each of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. The student-level effect sizes (ES_S) and slope effect sizes (ES_{Slope}) are presented in table A-10.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 10 days (including Saturdays, Sundays, and holidays).

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock-Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was no statistically detectable difference between groups in the covariate-adjusted spring pre-kindergarten or kindergarten means.

For the CMA-A Composite Score, there were no statistically significant differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

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

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each reading assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the TERA, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) the rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable t differences between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically detectable differences between groups on (a) the fall or (b) the spring pre-kindergarten assessments, but there was a statistically reliable difference on (c) the rate of growth from fall pre-kindergarten to spring pre-kindergarten (difference in rate of growth =12.9; ES_{Slope} = .5228, $p < .05$) favoring the treatment group. We did not obtain all three conditions necessary to indicate statistical evidence of a treatment effect on the WJ Spelling. On the ANCOVA, there was no statistically detectable difference between groups in the covariate-adjusted spring pre-kindergarten means.

For kindergarten, on the WJ Spelling test, there were no statistically detectable differences between groups in (d) covariate-adjusted means for the spring kindergarten assessments or in the (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was no statistically detectable difference between groups in the covariate-adjusted spring kindergarten means.

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

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted on the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

For the ANCOVA on the kindergarten CTOPP data, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

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

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatic Understanding subtest) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each language assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there were no statistically detectable differences between groups on (a) the fall assessment or (b) the spring pre-kindergarten assessment, but there was a statistically significant difference on the (c) rate of growth between groups from fall pre-kindergarten to spring pre-kindergarten (difference in rate of growth = -6.2; $ES_{Slope} = -.3262$, $p < .05$). In this instance, we do not have all three conditions necessary to indicate statistical evidence of a treatment effect on the PPVT. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for the spring pre-kindergarten assessment.

For kindergarten, on the PPVT, there was no statistically detectable difference in (d) covariate-adjusted means at the spring kindergarten or the (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was no statistically detectable difference in covariate-adjusted means for spring kindergarten.

On the TOLD Grammatic Understanding subtest, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses of the two language measures, we conclude that *Ladders to Literacy* did not have a statistically detectable effect on language development relative to the control condition.[5]

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (Social Skills Rating scale [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, we conducted an ANCOVA on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills, SSRS Problem Behaviors, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Social Skills scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for the spring pre-kindergarten or spring kindergarten assessments.

On the SSRS Problem Behaviors measure, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Problem Behaviors scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no

[5] The statistically significant difference between groups in rates of growth from pre-kindergarten spring to kindergarten spring does not "count" as a statistically significant test supporting a kindergarten effect because this slope does not address the impact of the intervention.

statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the PLBS, there were no statistically detectable differences in covariate-adjusted means on the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth between groups from fall to spring pre-kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring pre-kindergarten assessment.

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted, and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

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

Classroom Outcomes

The classroom-level effect sizes (ES_C) and slope effect sizes (ES_{Slope}) are presented in table A-10.

Overall classroom environment

We obtained observations on the Early Childhood Environment Rating Scale-Revised (ECERS-R) in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with the following covariates: teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or in (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained in the spring pre-kindergarten observation.

Based on the analyses of the ECERS-R, we conclude that *Creative Curriculum with Ladders to Literacy* did not have a statistically detectable effect on overall classroom quality.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year and conducted repeated measures analyses with the following covariates: teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site (note: no fall observation score was included). In addition, for each of the teacher-child relationship measures, ANCOVAs were conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the Arnett Detachment scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Harshness scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Permissiveness scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Positive Interactions scale, there were no statistically detectable differences in the covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

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

Classroom instruction

Because the classroom instruction measures (Teacher Behavior Rating Scale [TBRS] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in spring pre-kindergarten, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

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

The impact of *Creative Curriculum with Ladders to Literacy* on the child- and classroom-level measures is summarized in table A-10.

Table A-10. Secondary analysis results for *Creative Curriculum with Ladders to Literacy*

Measure	RM analysis start of treatment[1]	RM analysis Fall Pre-K	RM analysis Spring Pre-K	Fall-Spring slope	ANCOVA[2] Spring Pre-K	RM analysis kindergarten	Spring Pre-K-Spring K slope	ANCOVA kindergarten
				Student-level effect sizes				
Mathematics								
WJ Applied Problems	-.15	-.15	-.14	.0072	.03	-.33	-.0997	-.28
CMA-A Mathematics Composite	.11	.12	.18	.0510	.11	-.19	-.1960	-.28
Shape Composition[3]	-.08	-.07	.02	.0820	.10	-.10	-.0633	-.11
Reading								
TERA	.18	.09	-.30	-.3784	-.30	-.54	-.1271	-.60*
WJ Letter Word Identification	-.07	-.09	-.16	-.0734	.04	-.27	-.0585	-.17
WJ Spelling	-.36	-.24	.30	.5228*	.27	-.08	-.2009	-.23
Phonological awareness								
Pre-CTOPPP/CTOPP	.00	-.03	-.16	-.1262	-.12	†	†	-.10
Language								
PPVT	.03	-.04	-.38	-.3262*	-.22	-.30	.0438	-.29
TOLD	.04	-.01	-.22	-.2046	-.17	-.06	.0843	-.02
Behavior								
SSRS Social Skills	-.28	-.28	-.25	.0250	-.06	†	†	.17
SSRS Problem Behavior[4]	-.03	-.02	-.01	.0178	-.02	†	†	.02
PLBS/LBS	-.20	-.18	-.08	.0991	-.03	†	†	-.11
				Classroom-level effect sizes				
Global classroom quality								
ECERS-R	.86	.57	-.71	-1.2460	-.07	†	†	†
Teacher-child interaction								
Arnett Detachment[5]	-.42	-.24	.51	.7399	-.02	†	†	†
Arnett Harshness[5]	.85	.64	-.26	-.8805	-.07	†	†	†
Arnett Permissiveness[5]	.12	.29	1.02	.7151	.67	†	†	†
Arnett Positive Interactions	.67	.55	.03	-.5041	.99	†	†	†

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.
* $p < .05$
[1] The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).
[2] The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.
[3] Building Blocks, Shape Composition task
[4] Higher scores on this scale represent more negative child behaviors.
[5] Lower scores on this scale represent a more positive classroom environment.
NOTE: RM: Repeated Measures
ANCOVA: Analysis of covariance
Refer to the glossary for abbreviations of the measures.
SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Curiosity Corner:
Success for All Foundation (Kansas, Florida, and New Jersey sites)

We present analyses for each of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. The Success for All (SFA) team implemented its evaluation in three separate sites. Our discussion of the results focuses on the combined analyses of the three sites. The student-level effect sizes (ES_S) and slope effect sizes (ES_{Slope}) are presented in table A-11.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 14 days (including Saturdays, Sundays, and holidays) in Kansas, 35 days in New Jersey, and 49 days in Florida.

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock-Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the CMA-A Composite Score, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

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

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each reading assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the TERA, there were no statistically detectable differences between groups in the covariate adjusted means at the (a) fall pre-kindergarten assessment (b) spring pre-kindergarten assessment, or (c) the rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there was no statistically significant difference between groups in the covariate-adjusted spring pre-kindergarten means.

For kindergarten, on the TERA, there was (d) a statistically significant difference on the spring kindergarten assessment ($ES_S = .43$, $p < .05$) and (e) a statistically significant difference in the rate of growth from spring pre-kindergarten to spring kindergarten (difference in rate of growth = .05; $ES_{Slope} = .1771$, $p < .05$). On the ANCOVA, there was no statistically significant difference between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there were no statistically significant differences in covariate-adjusted means at the (a) fall, or (b) spring pre-kindergarten assessments, and (c) no difference in the rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there was no statistically detectable difference between groups in the covariate-adjusted spring pre-kindergarten means.

For kindergarten, on the WJ Letter Word Identification test, there was a statistically significant difference in the spring kindergarten assessment ($ES_S = .43$, $p < .05$) and a statistically significant difference in the rate of growth from spring pre-kindergarten to spring kindergarten (difference in rate of growth = 4.74; $ES_{Slope} = .1806$, $p < .05$). On the ANCOVA, there was no statistically significant difference between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically significant differences in covariate-adjusted means at the fall pre-kindergarten, spring pre-kindergarten, or kindergarten assessments, and no statistically significant differences in rates of growth from fall to spring pre-kindergarten and spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically significant differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically significant differences between groups in the covariate-adjusted spring kindergarten means.

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

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted of the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

For the ANCOVA on the kindergarten CTOPP data, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

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

Language assessments

Data from the two language measures (Peabody Picture and Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatic Understanding scale) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each language assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there were no statistically detectable t differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the TOLD Grammatic Understanding subtest, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA Grammatic Understanding scale, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

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

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (Social Skills Rating Scale [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, we conducted an ANCOVA on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Social Skills scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the SSRS Problem Behaviors measure, there was a statistically significant difference between groups on the (a) fall pre-kindergarten assessment ($ES_S = .53$, $p < .05$). There were no statistically detectable differences between groups on the (b) spring pre-kindergarten assessment, or the (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. The statistically reliable difference in Problem Behavior scores at the fall assessment suggests either the nonequivalence of treatment or control groups or an early treatment effect. To examine the possibility of an early treatment effect, we extrapolated back to the beginning of the school year and found a statistically reliable difference between groups on the Problem Behaviors measure ($ES_S = .56$, $p < .05$). This finding suggests, but does not prove, nonequivalence at the start of treatment. On the SSRS Problem Behaviors scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically significant differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the PLBS, there were no statistically detectable differences in covariate-adjusted means on the (a) fall pre-kindergarten assessment or (b) spring pre-kindergarten assessment, and no statistically detectable difference in (c) rate of growth between groups from fall to spring pre-kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring pre-kindergarten assessment.

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

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

Classroom Outcomes

The classroom-level effect sizes (ES_{Slope}) and slope effect sizes (ES_{Slope}) are presented in table A-11.

Overall classroom environment

We obtained observations on the Early Childhood Environment Rating Scale-Revised (ECERS-R) in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with the following covariates: teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R, there were no statistically detectable differences on the (a) covariate-adjusted, (b) spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

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

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year and conducted repeated measures analyses with the following covariates: teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site (note: no fall observation score was included). In addition, for each of the teacher-child relationship measures, ANCOVAs were conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the Arnett Detachment scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, there was no statistically significant difference between groups on the spring pre-kindergarten observation.

On the Arnett Harshness scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Permissiveness scale, there was a statistically significant difference between groups in the (a) covariate-adjusted means for the fall pre-kindergarten observation ($ES_C = -1.46$, $p < .05$). There was no statistically detectable difference on the (b) spring pre-kindergarten observations or the (c) rate of change from the fall to spring observation. The statistically reliable difference in Arnett Permissiveness scores at the fall observation suggests either the nonequivalence of treatment or control groups or early implementation of the study curriculum. To examine the possibility of an effect related to early implementation of the curriculum, we extrapolated back to the beginning of the school year and found a statistically reliable difference favoring the treatment group on the Permissiveness measure ($ES_C = -1.57$, $p < .05$). On the ANCOVA, no statistically significant difference was obtained on the spring pre-kindergarten observation.[6]

On the Arnett Positive Interactions scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

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

Classroom instruction

Because the classroom instruction measures (Teacher Behavior Rating Scale [TBRS] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in spring pre-kindergarten, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for *Curiosity Corner*

The impact of *Curiosity Corner* on the child- and classroom-level measures is summarized in table A-11.

[6] Even though there was a statistically significant difference between groups on the extrapolated start of treatment means, on the ANCOVA analysis, which covaries out any differences between groups at the fall observation, we did not obtain a statistically significant difference between groups on the spring pre-kindergarten observation.

Table A-11. Secondary analysis results for *Curiosity Corner*

Measure	RM analysis start of treatment[1]	RM analysis Fall Pre-K	RM analysis Spring Pre-K	Fall-Spring slope	ANCOVA[2] Spring Pre-K	RM analysis kindergarten	Spring Pre-K-Spring K slope	ANCOVA kindergarten
				Student-level effect sizes				
Mathematics								
WJ Applied Problems	.06	.06	.10	.0318	.10	.26	.0885	.06
CMA-A Mathematics Composite	.01	.01	.01	.0001	-.01	-.05	-.0330	.03
Shape Composition[3]	-.11	-.06	.16	.2143	.07	.32	.0901	.41
Reading								
TERA	.33	.29	.10	-.1816	-.06	.43*	.1771*	.32
WJ Letter Word Identification	.26	.23	.09	-.1328	-.02	.43*	.1806*	.29
WJ Spelling	-.16	-.12	.04	.1515	.05	.20	.0906	.19
Phonological awareness								
Pre-CTOPPP/CTOPP	-.06	-.01	.18	.1866	-.01	†	†	.25
Language								
PPVT	-.04	-.04	-.01	.0273	-.04	.14	.0785	.17
TOLD	-.02	-.03	-.08	-.0409	-.05	.15	.1198	.15
Behavior								
SSRS Social Skills	-.26	-.23	-.06	.1598	-.10	†	†	.32
SSRS Problem Behavior[4]	.56*	.53*	.43	-.1056	.07	†	†	-.08
PLBS/LBS	-.43	-.40	-.25	.1425	.02	†	†	.11
				Classroom-level effect sizes				
Global classroom quality								
ECERS-R	-.69	-.65	-.48	.1661	-.36	†	†	†
Teacher-child interaction								
Arnett Detachment[5]	-.07	-.14	-.41	-.2668	1.40	†	†	†
Arnett Harshness[5]	.34	.30	.14	-.1564	1.08	†	†	†
Arnett Permissiveness[5]	-1.57*	-1.46*	-.98	.4708	-.60	†	†	†
Arnett Positive Interactions	.72	.59	.02	-.5506	-1.43	†	†	†

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

* $p < .05$

[1] The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).

[2] The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.

[3] Building Blocks, Shape Composition task

[4] Higher scores on this scale represent more negative child behaviors.

[5] Lower scores on this scale represent a more positive classroom environment.

NOTE: RM: Repeated Measures
ANCOVA: Analysis of covariance

Refer to the glossary for abbreviations of the measures.

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

Doors to Discovery: University of Texas Health Science Center at Houston (Texas site)

We present analyses for each of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. The student-level effect sizes (ES_S) and slope effect sizes (ES_{Slope}) are presented in table A-12.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 20 days (including Saturdays, Sundays, and holidays).

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock-Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). Each model included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (fall assessment score was not included). For each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, or between groups in the covariate-adjusted spring kindergarten means.

For the CMA-A Composite Score, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

There was no statistically detectable difference between groups in (d) the spring kindergarten assessment, but there was a statistically reliable difference between groups in the (e) rates of growth from spring pre-kindergarten to spring kindergarten (difference in rates of growth = -.04; ES_{Slope} = -.1551, $p < .05$). On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

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

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included

the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each reading assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the TERA, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) the rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically detectable differences between groups in the covariate-adjusted means at the fall pre-kindergarten, spring pre-kindergarten, or spring kindergarten assessments, and no statistically detectable differences in rates of growth from fall to spring pre-kindergarten and spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

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

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted on the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

For the ANCOVA on the kindergarten CTOPP data, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

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

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatic Understanding subtest) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there were no statistically detectable differences in covariate-adjusted means at the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the TOLD Grammatic Understanding subtest, there (a) was a statistically detectable difference between groups on the fall assessment favoring the treatment group (ES_S = .38, $p < .05$), but no statistically detectable differences (b) at the spring pre-kindergarten assessment, or (c) in the rate of growth from fall pre-kindergarten to spring pre-kindergarten. A statistical difference at the fall assessment could reflect either the failure of randomization to create equivalent groups or an early treatment effect. Extrapolating back to the beginning of the school year, we found a statistically reliable difference between groups favoring the *Doors to Discovery* group (ES_S = .42, $p < .05$). This finding suggests, but does not prove, nonequivalence at the start of treatment. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten. There were no statistically detectable differences (d) at the spring kindergarten assessment or (e) in the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for the spring kindergarten assessment.

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

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, we conducted an ANCOVA on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences in covariate-adjusted means at (a) the fall pre-kindergarten or (b) spring pre-kindergarten assessments, and (c) no statistically detectable difference in the rate of growth between groups from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Social Skills subscale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the SSRS Problem Behaviors measure, there were no statistically detectable differences in covariate-adjusted means at (a) the fall pre-kindergarten or (b) spring pre-kindergarten assessments, and (c) no statistically detectable difference in the rate of growth between groups from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Problem Behaviors subscale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the Preschool Learning Behaviors Scale, there were no statistically detectable differences in covariate-adjusted means at (a) the fall pre-kindergarten or (b) spring pre-kindergarten assessments, and (c) no statistically detectable difference in the rate of growth between groups from fall to spring pre-kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring pre-kindergarten assessment.

On the Learning Behaviors Scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

Based on the analyses of the four behavioral measures, we conclude that *Doors to Discovery* did not have a statistically detectable effect on social and learning behaviors relative to the control condition.

Classroom Outcomes

The classroom-level effect sizes (ES_C) and slope effect sizes (ES_{Slope}) are presented in table A-12.

Overall classroom environment

We obtained observations on the ECERS-R in fall and spring of the pre-kindergarten year and conducted repeated measures analyses with the following covariates: teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically significant difference was obtained on the spring pre-kindergarten observation.

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

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, for each of the teacher-child relationships measures, ANCOVAs were conducted with the following covariates: *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site.

On the Arnett Detachment scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Harshness scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Permissiveness scale, there was a statistically significant difference between groups in the (a) covariate-adjusted means for the fall pre-kindergarten observation ($ES_C = 1.06$, $p < .05$). However, there were no statistically detectable differences between groups on the (b) spring pre-kindergarten observation, or

(c) the rate of change from the fall to spring observation. The statistically reliable difference in Arnett Permissiveness scores at the fall observation suggests either the nonequivalence of treatment or control groups or early implementation of the study curriculum. To examine the possibility of early implementation of the study curriculum, we extrapolated back to the beginning of the school year and found a statistically reliable difference favoring the treatment group on the Permissiveness measure ($ES_C = 1.28$, $p < .05$). *Doors to Discovery* teachers were more permissive in their interactions with students relative to teachers in the control classrooms as measured by the Arnett Permissiveness scale. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.[7]

On the Arnett Positive Interactions scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses of teacher-child relationship measures, we conclude that *Doors to Discovery* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

Because the classroom instruction measures (Teacher Behavior Rating Scale [TBRS] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in spring pre-kindergarten, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for *Doors to Discovery*

The impact of *Doors to Discovery* on the child- and classroom-level measures is summarized in table A-12.

[7] Even though there was a statistically significant difference between groups on the extrapolated start of treatment means, on the ANCOVA analysis, which covaries out any differences between groups at the fall observation, we did not obtain a statistically significant difference between groups on the spring pre-kindergarten observation.

Table A-12. Secondary analysis results for *Doors to Discovery*

Measure	RM analysis start of treatment[1]	RM analysis Fall Pre-K	RM analysis Spring Pre-K	Fall-Spring slope	ANCOVA[2] Spring Pre-K	RM analysis kindergarten	Spring Pre-K-Spring K slope	ANCOVA kindergarten
				Student-level effect sizes				
Mathematics								
WJ Applied Problems	.13	.11	.01	-.0932	.00	-.02	-.0185	-.05
CMA-A Mathematics Composite	.03	.05	.13	.0816	.16	-.16	-.1551*	-.15
Shape Composition[3]	-.11	-.11	-.13	-.0155	-.13	-.12	.0021	-.07
Reading								
TERA	.18	.16	.06	-.1004	-.05	-.05	-.0586	-.15
WJ Letter Word Identification	.10	.10	.10	-.0042	.09	-.09	-.0993	-.14
WJ Spelling	.16	.14	.06	-.0807	.04	-.12	-.0965	-.13
Phonological awareness								
Pre-CTOPPP/CTOPP	.15	.16	.18	.0245	.14	†	†	-.09
Language								
PPVT	.23	.21	.15	-.0666	.01	.18	.0196	.06
TOLD	.42*	.38*	.17	-.2026	.04	.06	-.0558	-.07
Behavior								
SSRS Social Skills	.03	-.01	-.18	-.1683	-.18	†	†	-.05
SSRS Problem Behavior[4]	-.39	-.34	-.14	.1997	.11	†	†	.46
PLBS/LBS	.12	.06	-.18	-.2367	-.26	†	†	-.32
				Classroom-level effect sizes				
Global classroom quality								
ECERS-R	.23	.26	.39	.1262	.19	†	†	†
Teacher-child interaction								
Arnett Detachment[5]	-.43	-.36	-.07	.2828	.11	†	†	†
Arnett Harshness[5]	-.34	-.35	-.38	-.0253	-.17	†	†	†
Arnett Permissiveness[5]	1.28*	1.06*	.13	-.9123	-.04	†	†	†
Arnett Positive Interactions	.40	.40	.38	-.0121	.15	†	†	†

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

* $p < .05$

[1] The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).

[2] The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.

[3] Building Blocks, Shape Composition task

[4] Higher scores on this scale represent more negative child behaviors.

[5] Lower scores on this scale represent a more positive classroom environment.

NOTE: RM: Repeated Measures
ANCOVA: Analysis of covariance

Refer to the glossary for abbreviations of the measures.

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

Let's Begin with the Letter People:
University of Texas Health Science Center at Houston (Texas site)

We present analyses for each of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. The student-level effect sizes (ES_S) and slope effect sizes (ES_{Slope}) are presented in table A-13.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 20 days (including Saturdays, Sundays, and holidays).

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (WJ Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means or in the covariate-adjusted spring kindergarten means.

For the CMA-A Composite Score, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were no statistically detectable differences between groups on the (a) fall assessment and (b) spring pre-kindergarten assessment, but there was a statistically reliable difference in the rates of growth from fall pre-kindergarten to spring pre-kindergarten (difference in rates of growth = .45; ES_{Slope} = .4783, $p < .01$) favoring the treatment group. Finally, there were no statistically detectable differences on (d) the spring kindergarten assessment and (e) the rates of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses of the three mathematics measures, we conclude that *Let's Begin with the Letter People* did not have an effect on mathematics development relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Language Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each reading assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the TERA, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rates of growth from fall pre-kindergarten to spring pre-kindergarten. There was no statistically significant difference between groups on (d) the spring kindergarten assessment or (e) the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically detectable differences in covariate-adjusted means at the fall pre-kindergarten, or spring kindergarten assessments, and no statistically detectable differences in rates of growth from fall to spring pre-kindergarten and spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses of the three reading measures, we conclude that *Let's Begin with the Letter People* did not have a statistically detectable effect on reading relative to the control condition.

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted on the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

For the ANCOVA on the kindergarten CTOPP data, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses of the phonological awareness measures, we conclude that *Let's Begin with the Letter People* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and TOLD Grammatic Understanding subtest) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the TOLD Grammatic Understanding subtest, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses of the two language measures, we conclude that *Let's Begin with the Letter People* did not have a statistically detectable effect on language development relative to the control condition.

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, we conducted an ANCOVA on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences in covariate-adjusted means at (a) the fall pre-kindergarten or (b) spring pre-kindergarten assessments, and (c) no statistically detectable difference in the rate of growth between groups from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Social Skills scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the SSRS Problem Behaviors measure, there were no statistically detectable differences in covariate-adjusted means at (a) the fall pre-kindergarten or (b) spring pre-kindergarten assessments, and (c) no statistically detectable difference in the rate of growth between groups from fall to spring pre-kindergarten. On the SSRS Problem Behaviors subscale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten.

On the PLBS, there were no statistically detectable differences in covariate-adjusted means at (a) the fall pre-kindergarten or (b) spring pre-kindergarten assessments, and (c) no statistically detectable difference in the rate of growth between groups from fall to spring pre-kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring pre-kindergarten assessment.

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

Based on the analyses of the behavior measures, we conclude that *Let's Begin with the Letter People* did not have a statistically detectable effect on social and learning behaviors relative to the control condition.

Classroom outcomes

The classroom-level effect sizes (ES_C) and slope effect sizes (ES_{Slope}) are presented in table A-13.

Overall classroom environment

We obtained observations on the Early Childhood Environment Rating Scale-Revised (ECERS-R) in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R, there were no statistically significant differences between groups on (a) the means from the fall pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. However, there was a statistically significant difference between groups (b) the spring pre-kindergarten observation (ES_C = .82, $p < .05$), such that treatment classrooms were rated as providing a more positive classroom environment. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses of the ECERS-R, we conclude that *Let's Begin with the Letter People* had no effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, for each of the teacher-child relationship measures, ANCOVAs were conducted with the following covariates: *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site.

On the Arnett Detachment scale, there were no statistically detectable differences between groups in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Harshness scale, there was no statistically detectable difference between groups on (a) the means from the fall pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. However, there was a statistically significant on difference between groups at the time of (b) the spring pre-kindergarten observation (ES_C = -.95, $p < .05$), such that relative to control group teachers, treatment group teachers were rated as exhibiting less irritation toward the children and being less likely to use threats to manage children's behaviors. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Permissiveness scale, there was a statistically reliable difference between groups in (a) the means from the fall pre-kindergarten observation (ES_C = .99, $p < .01$) and (c) the rate of change from the fall to spring observation (difference in rates of growth = -.35; ES_{Slope} = -1.016, $p < .05$). The statistically reliable difference in Arnett Permissiveness scores on the fall observation suggests either the nonequivalence of treatment or control groups or early implementation of the study curriculum. To examine the possibility of an effect related to early implementation of the curriculum, we extrapolated back to the beginning of the school year and found a statistically reliable difference favoring the treatment group on the Arnett Harshness

($ES_C = 1.23$, $p < .01$). However, no statistically detectable differences were found between groups on (b) the spring pre-kindergarten observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.[8]

On the Arnett Positive Interactions scale, there were no statistically detectable differences between groups in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses of the teacher-child relationship measures, we conclude that *Let's Begin with the Letter People* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

Because the classroom instruction measures (Teacher Behavior Rating Scale [TBRS] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in spring pre-kindergarten, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for *Let's Begin with the Letter People*

The impact of *Let's Begin with the Letter People* on the child- and classroom-level measures is summarized in table A-13.

[8] Even though there was a statistically significant difference between groups on the extrapolated start of treatment means, on the ANCOVA analysis, which covaries out any differences between groups at the fall observation, we did not obtain a statistically significant difference between groups on the spring pre-kindergarten observation.

Table A-13. Secondary analysis results for *Let's Begin with the Letter People*

Measure	RM analysis start of treatment[1]	RM analysis Fall Pre-K	RM analysis Spring Pre-K	Fall-Spring slope	ANCOVA[2] Spring Pre-K	RM analysis kindergarten	Spring Pre-K-Spring K slope	ANCOVA kindergarten
				Student-level effect sizes				
Mathematics								
WJ Applied Problems	-.08	-.09	-.10	-.0179	-.03	-.13	-.0151	-.10
CMA-A Mathematics Composite	.15	.15	.15	.0039	.12	-.07	-.1165	-.12
Shape Composition[3]	-.40*	-.28	.21	.4783**	.26	-.06	-.1427	-.00
Reading								
TERA	-.03	-.03	.02	.0411	.04	-.13	-.0766	-.12
WJ Letter Word Identification	-.16	-.11	.10	.2042	.19	-.18	-.1516	-.19
WJ Spelling	-.03	.01	.17	.1565	.15	-.06	-.1239	-.13
Phonological awareness								
Pre-CTOPPP/CTOPP	.12	.08	-.13	-.2006	-.16	†	†	-.13
Language								
PPVT	.14	.11	-.03	-.1348	-.08	.00	.0133	-.02
TOLD	.06	.07	.08	.0147	.08	-.12	-.1086	-.16
Behavior								
SSRS Social Skills	-.46	-.43	-.27	.1520	.02	†	†	.24
SSRS Problem Behavior[4]	-.13	-.12	-.06	.0578	.02	†	†	.06
PLBS/LBS	-.16	-.21	-.44	-.2154	-.35	†	†	-.10
				Classroom-level effect sizes				
Global classroom quality								
ECERS-R	-.07	.10	.82*	.7064	.74	†	†	†
Teacher-child interaction								
Arnett Detachment[5]	.22	.17	-.07	-.2316	-.15	†	†	†
Arnett Harshness[5]	.01	-.17	-.95*	-.7545	-.85	†	†	†
Arnett Permissiveness[5]	1.23**	.99**	-.05	-1.016*	-.29	†	†	†
Arnett Positive Interactions	.02	.11	.48	.3693	.37	†	†	†

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

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

[1] The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).

[2] The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.

[3] Building Blocks, Shape Composition task

[4] Higher scores on this scale represent more negative child behaviors.

[5] Lower scores on this scale represent a more positive classroom environment.

NOTE: RM: Repeated Measures
ANCOVA: Analysis of covariance

Refer to the glossary for abbreviations of the measures.

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

Early Literacy and Learning Model (ELLM): University of North Florida (Florida-UNF site)

We present analyses for each of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. Within each domain, we present the repeated measures models followed by the ANCOVAs that included the fall assessment as one of the covariates in the model. The University of North Florida (Florida-UNF) research team implemented its evaluation in three separate sites; table A-14 presents results for the combined analysis. Our discussion of the results focuses on the combined analysis of the three sites. The student-level effect sizes (ES_S) and slope effect sizes (ES_{Slope}) are presented in table A-14.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 28 days (including Saturdays, Sundays, and holidays) in County A, 27 days in County B, and 21 days in County C.

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (WJ Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was no statistically detectable difference between groups in the covariate-adjusted spring pre-kindergarten means or the covariate-adjusted spring kindergarten means.

For the CMA-A Composite Score, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable difference between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable difference between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the three mathematics measures, we conclude that the *Early Literacy and Learning Model* did not have a statistically detectable effect on mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included

the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each reading assessment, and ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the TERA, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable difference between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable difference between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically detectable differences between groups on the (a) fall assessment and (b) the spring pre-kindergarten assessment, but there was (c) a statistically reliable difference between groups in the rate of growth from fall pre-kindergarten to spring pre-kindergarten (difference in rate of growth = 8.54; $ES_{Slope} = .3179$, $p < .01$), such that children in the treatment group learned at a faster rate than children in the control group. In this instance, we do not have all three conditions necessary to indicate statistical evidence of a treatment effect on the WJ Spelling test. There were no statistically detectable differences (d) on the spring kindergarten assessment or (e) in the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable difference between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the three reading measures, we conclude that the *Early Literacy and Learning Model* did not have a statistically detectable effect on reading relative to the control condition.

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted on the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups in the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

For the ANCOVA on the kindergarten CTOPP data, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses of the phonological awareness measures, we conclude that the *Early Literacy and Learning Model* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatic Understanding subtest) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and, and mother's education (note: no fall assessment score was included). In addition, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. There was a (d) statistically significant difference between groups on the spring kindergarten assessment (ES_S = .34, $p < .05$) favoring the treatment group, but there was no statistically detectable difference (e) in rates of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted for spring pre-kindergarten or spring kindergarten assessments.

On the TOLD Grammatic Understanding subtest, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. There was a (d) statistically significant difference between groups on the spring kindergarten assessment (ES_S = .44, $p < .05$) favoring the treatment group, but there was no statistically detectable difference in (e) the rates of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten, but there was a significant difference in the covariate-adjusted means for spring kindergarten (ES_S = .39, $p < .01$).

Based on the analyses for the two language measures, we conclude that the *ELLM* did not have a statistically detectable effect on language development relative to the control condition in pre-kindergarten but had a positive effect in kindergarten.

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, ANCOVA analyses were conducted on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences between groups on the (a) fall pre-kindergarten or (b) spring pre-kindergarten assessments, and (c) no statistically detectable difference in the rate of growth between groups from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Social Skills scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the SSRS Problem Behaviors measures, there were no statistically detectable differences between groups on (a) the means from the fall pre-kindergarten assessment, (b) the spring pre-kindergarten assessment, or the (c) rate of change from the fall to spring pre-kindergarten. On the SSRS Problem Behaviors scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten.

On the PLBS, there were no statistically detectable differences in covariate-adjusted means at (a) the fall pre-kindergarten or (b) spring pre-kindergarten assessments, and (c) no statistically detectable difference in the rate of growth between groups from fall to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten assessment.

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

Based on the analyses of the behavioral measures, we conclude that the *Early Literacy and Learning Model* did not have a statistically detectable effect on social and learning behaviors relative to the control condition.

Classroom Outcomes

The classroom-level effect sizes (ES_C) and slope effect sizes (ES_{Slope}) are presented in table A-14.

Overall classroom environment

We obtained observations on the Early Childhood Environment Rating Scale-Revised (ECERS-R) in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R, there were no statistically detectable difference between groups in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) the spring pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

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

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, for each of the teacher-child relationship measures, ANCOVAs were conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the Arnett Detachment scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Harshness scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Permissiveness scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Positive Interactions scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses for the four teacher-child relationship measures, we conclude that the *ELLM* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

Because the classroom instruction measures (Teacher Behavior Rating Scale [TBRS] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in spring pre-kindergarten, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for the *Early Literacy and Learning Model*

The impact of *Early Literacy and Learning Model* on the child- and classroom-level measures is summarized in table A-14.

Table A-14. Secondary analysis results for *Early Literacy and Learning Model*

Measure	RM analysis start of treatment[1]	RM analysis Fall Pre-K	RM analysis Spring Pre-K	Fall-Spring slope	ANCOVA[2] Spring Pre-K	RM analysis kindergarten	Spring Pre-K-Spring K slope	ANCOVA kindergarten
				Student-level effect sizes				
Mathematics								
WJ Applied Problems	.06	.06	.10	.0318	-.01	.26	.0885	.28
CMA-A Mathematics Composite	.01	.01	.01	.0001	.01	-.05	-.0330	-.09
Shape Composition[3]	.17	.11	-.14	-.2432	-.19	.03	.0881	-.02
Reading								
TERA	.12	.13	.15	.0217	.04	.30	.0788	.20
WJ Letter Word Identification	-.14	-.12	-.05	.0713	-.04	.00	.0260	.03
WJ Spelling	-.29	-.21	.11	.3179**	.21	.04	-.0368	.08
Phonological awareness								
Pre-CTOPPP/CTOPP	-.06	-.01	.18	.1866	.14	†	†	.08
Language								
PPVT	.28	.26	.17	-.0843	-.06	.34*	.0891	.12
TOLD	-.01	.02	.15	.1328	.05	.44*	.1531	.39**
Behavior								
SSRS Social Skills	-.26	-.23	-.06	.1598	.08	†	†	.27
SSRS Problem Behavior[4]	-.38*	-.35	-.24	.1073	-.01	†	†	.23
PLBS/LBS	.05	.07	.14	.0722	.07	†	†	.04
				Classroom-level effect sizes				
Global classroom quality								
ECERS-R	-.69	-.65	-.48	.1661	-.14	†	†	†
Teacher-child interaction								
Arnett Detachment[5]	-.07	-.14	-.41	-.2668	-.02	†	†	†
Arnett Harshness[5]	.08	-.01	-.40	-.3783	.02	†	†	†
Arnett Permissiveness[5]	.09	.03	-.24	-.2666	-.10	†	†	†
Arnett Positive Interactions	-.53	-.38	.29	.6519	-.01	†	†	†

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

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

[1] The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).

[2] The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.

[3] Building Blocks, Shape Composition task

[4] Higher scores on this scale represent more negative child behaviors.

[5] Lower scores on this scale represent a more positive classroom environment.

NOTE: RM: Repeated Measures
ANCOVA: Analysis of covariance

Refer to the glossary for abbreviations of the measures.

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

Language-Focused Curriculum: University of Virginia (Virginia site)

We present analyses for each of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. Within each domain, we present the repeated measures models followed by the ANCOVAs that included the fall assessment as one of the covariates in the model. The student-level effect sizes (ES_S) and slope effect sizes (ES_{Slope}) are presented in table A-15.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 28 days (including Saturdays, Sundays, and holidays).

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (WJ Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups in the (a) fall assessment, (b) spring pre-kindergarten assessment, but there was (c) a statistically significant difference between groups in the rate of growth from fall pre-kindergarten to spring pre-kindergarten (difference in rate of growth = 4.91; $ES_{Slope} = .2943$, $p < .05$). There were no statistically detectable differences (d) on the spring kindergarten assessment and (e) in the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the CMA-A Composite Score, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

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

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each reading assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the TERA, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable between groups in the covariate-adjusted spring kindergarten means.

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

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted on the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

For the ANCOVA on the kindergarten CTOPP data, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the phonological awareness measures, we conclude that the *Language-Focused Curriculum* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and TOLD Grammatic Understanding subtest) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the TOLD Grammatic Understanding subtest, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

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

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, we conducted an ANCOVA on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences between groups on the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or (c) the rate of growth from fall pre-kindergarten to spring pre-kindergarten On the SSRS Social Skills scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was e no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten. There was no statistically significant difference in the covariate-adjusted means for the spring kindergarten assessment.

On the SSRS Problem Behaviors measure, there were no statistically detectable differences between groups on the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Problem Behaviors scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten.

On the PLBS, there were no statistically detectable differences in covariate-adjusted means on the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth between groups from fall to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten assessment.

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

Based on the analyses of the behavioral measures, we conclude that the *Language-Focused Curriculum* did not have a statistically detectable effect on social and learning behaviors relative to the control condition.

Classroom Outcomes

The classroom-level effect sizes (ES_C) and slope effect sizes (ES_{Slope}) are presented in table A-15.

Overall classroom environment and teacher-child relationships

We obtained observations on the Early Childhood Environment Rating Scale-Revised (ECERS-R) in the fall and spring of the pre-kindergarten year. We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year. We did not conduct analyses using the ECERS and Arnett data for this site because of data integrity concerns. During the baseline data collection, one observer completed the observational ratings in 8 of the 12 classrooms at this research site. It was later determined that the ECERS-R and Arnett ratings from these eight classrooms were inflated. Due to concerns with the integrity of the data from these eight classrooms, the decision was made to exclude the classroom quality and teacher-child relationships data for this site from the report.

Classroom instruction

Because the classroom instruction measures (Teacher Behavior Rating Scale [TBRS] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in spring pre-kindergarten, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for *Language-Focused Curriculum*

The impact of *Language-Focused Curriculum* on the child- and classroom-level measures is summarized in table A-15.

Table A-15. Secondary analysis results for *Language-Focused Curriculum*

Measure	RM analysis start of treatment[1]	RM analysis Fall Pre-K	RM analysis Spring Pre-K	Fall-Spring slope	ANCOVA[2] Spring Pre-K	RM analysis kindergarten	Spring Pre-K-Spring K slope	ANCOVA kindergarten
				Student-level effect sizes				
Mathematics								
WJ Applied Problems	-.18	-.11	.20	.2943*	.27	.11	-.0464	.12
CMA-A Mathematics Composite	-.02	.00	.08	.0824	.12	.00	-.0423	.02
Shape Composition[3]	.08	.08	.08	.0055	-.01	.06	-.0126	.03
Reading								
TERA	.04	.07	.16	.0955	.09	.05	-.0617	-.07
WJ Letter Word Identification	.09	.09	.11	.0228	.11	.02	-.0509	-.05
WJ Spelling	.07	.10	.25	.1416	.29	.11	-.0722	.06
Phonological awareness								
Pre-CTOPPP/CTOPP	-.12	-.06	.20	.2565	.28	†	†	.03
Language								
PPVT	-.16	-.12	.02	.1423	.12	-.09	-.0612	-.03
TOLD	-.12	-.10	.01	.1019	-.04	-.07	-.0393	-.03
Behavior								
SSRS Social Skills	.20	.08	-.42	-.4904	-.51	†	†	-.07
SSRS Problem Behavior[4]	.40	.39	.37	-.0173	.21	†	†	-.05
PLBS/LBS	-.28	-.28	-.27	.0096	-.25	†	†	.10
				Classroom-level effect sizes				
Global classroom quality								
ECERS-R	—	—	—	—	—	†	†	†
Teacher-child interaction								
Arnett Detachment[5]	—	—	—	—	—	†	†	†
Arnett Harshness[5]	—	—	—	—	—	†	†	†
Arnett Permissiveness[5]	—	—	—	—	—	†	†	†
Arnett Positive Interactions	—	—	—	—	—	†	†	†

— Not available. Data were collected but not reported.
† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.
* $p < .05$
[1] The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).
[2] The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.
[3] Building Blocks, Shape Composition task
[4] Higher scores on this scale represent more negative child behaviors.
[5] Lower scores on this scale represent a more positive classroom environment.
NOTE: RM: Repeated Measures
ANCOVA: Analysis of covariance
For the *Language-Focused Curriculum*, we did not conduct analyses using the ECERS and Arnett data because of unreliable data. During the baseline data collection, one observer completed the observational ratings in eight of the 12 classrooms at this research site. It was later determined that the ECERS-R and Arnett ratings from these eight classrooms were inflated. Due to concerns with the integrity of the data from these eight classrooms, the decision was made to exclude the classroom quality and teacher-child relationships data for this site from the report. Refer to the glossary for abbreviations of the measures.
SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Literacy Express: Florida State University (Florida-FSU site)

We present analyses for each of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. Within each domain, we present the repeated measures models followed by the ANCOVAs that included the fall assessment as one of the covariates in the model. The student-level effect sizes (ES_S) and slope effect sizes (ES_{Slope}) are presented in table A-16.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 42 days (including Saturdays, Sundays, and holidays).

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock-Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the CMA-A Composite Score, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

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

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each reading assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the TERA, there were no statistically detectable differences between groups on the (a) fall assessment and (b) spring pre-kindergarten assessment, but there was significant difference between groups in (c) the rate of growth from fall pre-kindergarten to spring pre-kindergarten (difference in rate of growth = 2.10; $ES_{Slope} = .2815$, $p < .05$). There were no statistically detectable differences on the (d) spring kindergarten assessment and (e) no statistically detectable difference between groups in the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there was (a) a significant difference between groups on the fall assessment ($ES_S = .44$, $p < .05$). This result could indicate either an early treatment effect or failure of random assignment to produce equivalent groups. We extrapolated back to the start of the school year and found a statistically reliable difference in means ($ES_S = .47$, $p < .05$) at the start of the year. This difference suggests, but does not prove, nonequivalence of treatment and control groups. Because there was no evidence on any other measure of nonequivalence between groups at the start of treatment, we considered the groups to be equivalent. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

In the spring of pre-kindergarten, there were no statistically detectable differences between groups on (b) the covariate-adjusted means at the spring pre-kindergarten assessment and (c) the rates of growth fall to spring of the pre-kindergarten year. There were no statistically detectable differences on the (d) spring kindergarten assessment or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically detectable differences in covariate-adjusted means at the fall pre-kindergarten, spring pre-kindergarten, or spring kindergarten assessments, and no statistically detectable differences in rates of growth from fall to spring pre-kindergarten and spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

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

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted on the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups in the (a) fall pre-kindergarten assessment, or (b) spring pre-kindergarten assessment. There was, however, a statistically significant difference in (c) the rate of growth from fall pre-kindergarten to spring pre-kindergarten (differences in rates of growth = 1.35; $ES_{Slope} = .3217$, $p < .05$). On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

For the ANCOVA on the kindergarten CTOPP data, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

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

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and TOLD Grammatic Understanding subtest) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there were no statistically detectable differences between groups in the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the TOLD Grammatic Understanding subtest, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

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

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, we conducted an ANCOVA on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences between groups on the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or on (c) the rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the SSRS Problem Behaviors measure, there were no statistically detectable differences between groups on (a) the means from the fall pre-kindergarten, (b) the spring pre-kindergarten assessment, or the (c) rate of change from the fall to spring assessment. On the SSRS Problem Behaviors scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the PLBS, there were no statistically detectable differences in covariate-adjusted means on the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth between groups from fall to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten.

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted. On the ANCOVA, there was a significant difference in the covariate-adjusted means for spring kindergarten ($ES_S = -.38, p < .05$), such that children in the treatment group showed weaker learning behaviors than children in the control group.

Based on the analyses of the behavioral measures, we conclude that *Literacy Express* did not have a statistically detectable effect on social and learning behaviors relative to the control condition at spring of pre-kindergarten, but did have a negative effect on social and learning behaviors relative to the control condition at spring of kindergarten.

Classroom Outcomes

The classroom-level effect sizes (ES_C) and slope effect sizes (ES_{Slope}) are presented in table A-16.

Overall classroom environment

We obtained observations on the Early Childhood Environment Rating Scale-Revised (ECERS-R) in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, or (c) the rate of change from the fall to spring observation. There was a statistically significant difference between groups on the (b) spring pre-kindergarten observation ($ES_C = 1.29$, $p < .05$). On the ANCOVA, a statistically significant difference was obtained on the spring pre-kindergarten observation ($ES_C = 1.22, p < .05$) favoring the treatment group.

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

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, for each of the teacher-child relationship measures, ANCOVAs were conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the Arnett Detachment scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Harshness scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Permissiveness scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Positive Interactions scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses for the four teacher-child relationship measures, we conclude that *Literacy Express* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

Because the classroom instruction measures (Teacher Behavior Rating Scale [TBRS] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in spring pre-kindergarten, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for *Literacy Express*

The impact of *Literacy Express* on the child- and classroom-level measures is summarized in table A-16.

Table A-16. Secondary analysis results for *Literacy Express*

Measure	RM analysis start of treatment[1]	RM analysis Fall Pre-K	RM analysis Spring Pre-K	Fall-Spring slope	ANCOVA[2] Spring Pre-K	RM analysis kindergarten	Spring Pre-K-Spring K slope	ANCOVA kindergarten
				Student-level effect sizes				
Mathematics								
WJ Applied Problems	.02	.03	.05	.0260	-.04	-.02	-.0379	-.09
CMA-A Mathematics Composite	-.17	-.15	-.02	.1192	.03	-.21	-.0975	-.17
Shape Composition[3]	-.15	-.12	-.01	.1057	.05	-.14	-.0664	-.07
Reading								
TERA	-.19	-.12	.17	.2815*	.23	-.11	-.1518	-.02
WJ Letter Word Identification	.47*	.44*	.30	-.1341	-.02	.08	-.1154	-.29
WJ Spelling	.14	.12	.05	-.0671	-.07	.06	.0037	.06
Phonological awareness								
Pre-CTOPPP/CTOPP	-.26	-.19	.14	.3217*	.15	†	†	.08
Language								
PPVT	.05	.07	.17	.0963	.06	.16	-.0041	.09
TOLD	-.25	-.21	-.04	.1687	.04	.10	.0743	.13
Behavior								
SSRS Social Skills	.44	.35	-.06	-.4022	-.21	†	†	-.37
SSRS Problem Behavior[4]	-.60*	-.54	-.31	.2279	.00	†	†	.22
PLBS/LBS	.42	.38	.17	-.2053	-.02	†	†	-.38*
				Classroom-level effect sizes				
Global classroom quality								
ECERS-R	-.14	.12	1.29*	1.1353	1.22*	†	†	†
Teacher-child interaction								
Arnett Detachment[5]	.46	.17	-1.09	-1.230	-1.17	†	†	†
Arnett Harshness[5]	.03	-.13	-.84	-.6959	-.94	†	†	†
Arnett Permissiveness[5]	-1.12	-.82	.51	1.2987	.61	†	†	†
Arnett Positive Interactions	-.89	-.62	.56	1.1518	1.04	†	†	†

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.
* $p < .05$
[1] The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).
[2] The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.
[3] Building Blocks, Shape Composition task
[4] Higher scores on this scale represent more negative child behaviors.
[5] Lower scores on this scale represent a more positive classroom environment.
NOTE: RM: Repeated Measures
ANCOVA: Analysis of covariance
Refer to the glossary for abbreviations of the measures.
SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

DLM Early Childhood Express supplemented with Open Court Reading Pre-K:
Florida State University (Florida-FSU site)

We present analyses for each of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. Within each domain, we present the repeated measures models followed by the ANCOVAs that included the fall assessment as one of the covariates in the model. The student-level effect sizes (ES_S) and slope effect sizes (ES_{Slope}) are presented in table A-17.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 28 days (including Saturdays, Sundays, and holidays).

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock-Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment or (b) the rates of growth from fall pre-kindergarten to spring pre-kindergarten, but there was a statistically significant difference on the (c) spring pre-kindergarten assessment ($ES_S = .36$, $p < .01$). In this instance, we do not have all three conditions necessary to indicate statistical evidence of a treatment effect on the WJ Applied Problems. There was a statistically significant difference on the (d) spring kindergarten assessment ($ES_S = .48$, $p < .001$), but no statistically significant difference on (e) the rates of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically significant differences between groups in the covariate-adjusted spring pre-kindergarten means, but there was a statistically significant difference between groups in the covariate-adjusted spring kindergarten means ($ES_S = .31$, $p < .05$). We conclude there was no effect of *DLM Early Childhood Express supplemented with Open Court Reading Pre-K)* on the WJ Applied Problems for the spring pre-kindergarten assessment; however, *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* had a positive effect on the WJ Applied Problems for the spring of kindergarten assessment.

For the CMA-A Composite Score, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Across the three math measures, we did not obtain a consistent pattern of results and concluded that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K*did not have an impact on the mathematics outcome relative to the control condition. Although we did not obtain consistent evidence across the mathematics measures, the analyses of the WJ Applied Problems are consistent with the pattern of results from the reading and phonological awareness data. We examined the intercorrelations among measures for the combined control group (i.e., children in the control condition across all research/grantee sites) for the fall assessment. Scores on the WJ Applied Problems were moderately correlated with each of the reading, phonological awareness, and language measures (TERA: r =.60; WJ Letter Word Identification: r = .49, $p < .0001$; WJ Spelling: $r = .46, p < .0001$; Pre-CTOPPP: $r = .48, p < .0001$; PPVT: $r = .63, p < .0001$; TOLD Grammatic Understanding subtest: $r = .47, p < .0001$).

Based on the analyses of the three mathematics measures, we conclude that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* did not have a statistically detectable effect on mathematics relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each reading assessment, an ANCOVA was conducted in which the covariates were: fall assessment score, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the TERA, there was no statistically detectable difference between groups on the (a) fall assessment, but there were statistically reliable differences on the (b) spring pre-kindergarten assessment ($ES_S = .68, p < .001$) and (c) rates of growth from fall pre-kindergarten to spring pre-kindergarten (difference in rates of growth = 3.22; $ES_{Slope} = .4052, p < .001$). Taken together these three results provide statistical evidence of a treatment effect of *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* on the TERA. On the ANCOVA, there was a statistically significant difference between groups in the covariate-adjusted spring pre-kindergarten means ($ES_S = .40, p < .05$).

In addition, on the repeated measures analysis of the TERA, there was a statistically reliable difference between groups on the (d) spring kindergarten assessment ($ES_S = .76, p < .01$); there were no differences in the (e) rates of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was no statistically detectable difference between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there were differences between groups on the (a) fall assessment ($ES_S = .41, p < .05$) and (b) spring pre-kindergarten assessment ($ES_S = .51, p < .01$), but not in the (c) rates of growth from fall pre-kindergarten to spring pre-kindergarten. The statistically reliable difference in WJ Letter Word Identification scores at baseline suggests either the nonequivalence of treatment and control groups or an early treatment effect. To examine the possibility of an early treatment effect, we extrapolated back to the beginning of the school year and found a statistically reliable difference favoring the *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* group on the WJ Letter Word Identification test ($ES_S = .39, p < .05$). For the following reasons, this difference suggests, but does not prove, nonequivalence of treatment and control groups. The extrapolation is based on the average score for each group at the fall assessment and the rate of growth from fall to spring for each group. Although the slope from fall to spring is our best estimate of the rate of growth from the beginning of the school year to the fall assessment, if there were an early treatment effect, there is no theoretical reason to assume that growth was constant from the start of the school year to the spring assessment. The WJ Letter Word Identification test has a high floor—that is, there are few items for younger children and one additional correct answer can result in substantial differences in the standardized scores. For example, for a child age 4 years and 6 months, three correct answers yield a standardized score of 89 and four correct answers a score of 95. The fall assessment average standardized raw score for the control group was 91.53; their spring standardized raw score was 95.60.[9] The difference between the fall and spring assessments is less than one correct answer. Because the initial items on the test are letter identification items, curricula that focus first on learning letters have the possibility to produce a rapid early treatment effect.

In the spring of pre-kindergarten, statistically reliable differences were obtained on the WJ Letter Word Identification test ($ES_S = .51$, $p < .01$); however there was no statistically detectable difference in slopes from fall to spring of the pre-kindergarten year. Consequently the difference obtained at spring could simply reflect the difference obtained at baseline. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

For the spring kindergarten assessment, statistically reliable differences were obtained on the WJ Letter Word Identification test ($ES_S = .50$, $p < .01$); however, there was no statistically detectable difference in rates of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically detectable differences between groups on the (a) fall assessment or in (b) the rates of growth from fall pre-kindergarten to spring pre-kindergarten. However, there was a statistically reliable difference on the (c) spring pre-kindergarten assessment ($ES_S = .46$, $p < .01$). Taken together, these results do not meet the criteria for establishing statistical evidence of a treatment effect for a single outcome measure in the pre-kindergarten year. On the ANCOVA, there was a statistically significant difference between groups in the covariate-adjusted spring pre-kindergarten means ($ES_S = .27$, $p < .05$).

On the repeated measures analysis of the WJ Spelling, there were no statistically significant group differences on the (d) spring kindergarten assessment, or in the (e) rates of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there was no statistically significant difference between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses of the three reading measures, we conclude that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* had a positive effect on reading relative to the control condition in pre-kindergarten.

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted on the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Processing (CTOPP), Elision subtest data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups on the (a) pre-kindergarten fall assessment, or the (b) the rate of growth from fall pre-kindergarten to spring pre-kindergarten, but a statistically reliable difference between groups was found at the spring pre-kindergarten assessment ($ES_S = .32$, $p < .05$). On the ANCOVA for the Pre-CTOPPP, there was no statistically detectable difference between groups in the covariate-adjusted spring pre-kindergarten means.

[9] Schrank, F.A. and Woodcock, R.W. (2001). WJ III Compuscore and Profiles Program [Computer software]. *Woodcock-Johnson III*. Itasca, IL: Riverside Publishing.

For the ANCOVA on the kindergarten CTOPP data, there was a statistically significant difference between groups in the covariate-adjusted spring kindergarten means ($ES_S = 38, p < .05$).

Based on the analyses of the phonological awareness measures, we conclude that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* had no effect on phonological awareness relative to the control condition in pre-kindergarten and a positive effect in kindergarten.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatic Understanding subtest) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there was no statistically detectable difference between groups on (a) the fall assessment. In the spring of the pre-kindergarten year, there was a statistically reliable mean difference between groups ($ES_S = .40, p < .05$); however, there was (c) no statistically reliable difference between groups in the slopes. On the ANCOVA for the PPVT, there were no statistically significant differences in covariate-adjusted means for the spring pre-kindergarten assessment.

In spring of the kindergarten year, there was a statistically reliable mean difference between groups on the PPVT ($ES_S = .48, p < .01$); there was no statistically significant difference in the rates of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA analysis for the PPVT, there was no statistically significant difference in covariate-adjusted means for the spring kindergarten assessment.

For the TOLD Grammatic Understanding subtest, there was (a) a statistically reliable difference in means favoring children in the *Open Court Reading Pre-K* condition on the fall assessment ($ES_S = .38, p < .05$). We extrapolated scores back to the beginning of the year and found no statistically detectable differences at the start of the school year between the two groups. This result suggests that differences at the fall assessment reflect early treatment effects. In the spring of the pre-kindergarten year, there was (b) a statistically reliable mean difference in scores on the TOLD Grammatic Understanding subtest ($ES_S = .40, p < .01$), but there was (c) no statistically detectable difference between groups in the rates of growth from fall to spring pre-kindergarten. On the ANCOVA for the TOLD Grammatic Understanding subtest, there were no statistically detectable differences in covariate-adjusted means for the spring pre-kindergarten assessment.

In spring of the kindergarten year, there was a statistically reliable difference in means on the TOLD favoring children in the *Open Court* condition ($ES_S = .46, p < .01$). There was no statistically detectable difference between groups in the rates of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA analyses for TOLD Grammatic Understanding, there was no statistically detectable difference between groups in covariate-adjusted means for the spring kindergarten assessment.

Based on the analyses of the two language measures, we conclude that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* did not have an effect on language development relative to the control condition at spring of pre-kindergarten. However, *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* had a positive effect on language development relative to the control condition by spring of kindergarten.

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, we conducted an ANCOVA on the pre-kindergarten (SSRS

Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences between groups on the (a) fall pre-kindergarten assessment, or (b) spring pre-kindergarten assessment, but a statistically reliable difference between groups was found on (c) the rate of growth from fall pre-kindergarten to spring pre-kindergarten (difference in rates of growth = -6.4, ES_{Slope} = -.4253, $p < .05$). On the ANCOVA, there were no differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten.

On the SSRS Problem Behaviors measure, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, or (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten.

On the PLBS, there were no statistically detectable differences in covariate-adjusted means on the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth between groups from fall to spring pre-kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring pre-kindergarten assessment.

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

Based on the analyses of the four behavioral measures, we conclude that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* did not have a statistically detectable effect on social and learning behaviors relative to the control condition.

Classroom Outcomes

The classroom-level effect sizes (ES_C) and slope effect sizes (ES_{Slope}) are presented in table A-17.

Overall classroom environment

We obtained observations on the Early Childhood Environment Rating Scale-Revised (ECERS-R) in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with the following covariates: teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses of the ECERS-R, we conclude that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* did not have a statistically detectable effect on overall classroom environment relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, for each of the teacher-child relationship measures, ANCOVAs were conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the Arnett Detachment scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Harshness scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, there was no statistically detectable difference between groups on the spring pre-kindergarten.

On the Arnett Permissiveness scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Positive Interactions scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses of the teacher-child relationship measures, we conclude that *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

Because the classroom instruction measures (Teacher Behavior Rating Scale [TBRS] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in spring pre-kindergarten, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for *DLM Early Childhood Express supplemented with Open Court Reading Pre-K*

The impact of *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* on the child- and classroom-level measures is summarized in table A-17.

Table A-17. Secondary analysis results for *DLM Early Childhood Express supplemented with Open Court Reading Pre-K*

Measure	RM analysis start of treatment[1]	RM Analysis Fall Pre-K	RM analysis Spring Pre-K	Fall-Spring slope	ANCOVA[2] Spring Pre-K	RM analysis kindergarten	Spring Pre-K-Spring K slope	ANCOVA kindergarten
				Student-level effect sizes				
Mathematics								
WJ Applied Problems	.17	.21	.36**	.1497	.19	.48***	.0608	.31*
CMA-A Mathematics Composite	-.14	-.08	.17	.2526	.22	.13	-.0232	.10
Shape Composition[3]	-.16	-.08	.24	.3154	.21	.09	-.0806	.10
Reading								
TERA	.17	.27	.68***	.4052***	.40*	.76**	.0418	.33
WJ Letter Word Identification	.39*	.41*	.51**	.0945	.23	.50**	-.0070	.13
WJ Spelling	.21	.26	.46**	.1984	.27*	.22	-.1282	.14
Phonological awareness								
Pre-CTOPPP/CTOPP	.02	.08	.32*	.2377	.25	†	†	.38*
Language								
PPVT	.31	.33	.40*	.0754	.15	.48**	.0408	.22
TOLD	.38	.38*	.40**	.0207	.26	.46**	.0329	.27
Behavior								
SSRS Social Skills	.42	.32	-.11	-.4253*	-.28	†	†	-.18
SSRS Problem Behavior[4]	-.04	-.01	.11	.1247	.11	†	†	.01
PLBS/LBS	-.08	-.09	-.16	-.0671	-.08	†	†	-.13
				Classroom-level effect sizes				
Global classroom quality								
ECERS-R	-.21	-.11	.34	.4320	.55	†	†	†
Teacher-child interaction								
Arnett Detachment[5]	.06	.04	-.06	-.0928	-.22	†	†	†
Arnett Harshness[5]	-.37	-.43	-.70	-.2574	-.85	†	†	†
Arnett Permissiveness[5]	-.37	-.29	.05	.3327	.23	†	†	†
Arnett Positive Interactions	-.66	-.46	.43	.8638	.53	†	†	†

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.
* $p < .05$; ** $p < .01$; *** $p < .001$
[1] The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).
[2] The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.
[3] Building Blocks, Shape Composition task
[4] Higher scores on this scale represent more negative child behaviors.
[5] Lower scores on this scale represent a more positive classroom environment.
NOTE: RM: Repeated Measures
ANCOVA: Analysis of covariance
Refer to the glossary for abbreviations of the measures.
SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Pre-K Mathematics supplemented with DLM Early Childhood Express Math software: University of California, Berkeley and University at Buffalo, State University of New York (California and New York sites)

We present analyses for each of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. Within each domain, we present the repeated measures models followed by the ANCOVAs that included the fall assessment as one of the covariates in the model. The student-level effect sizes (ES_S) and slope effect sizes (ES_{Slope}) are presented in table A-18.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 19 days (including Saturdays, Sundays, and holidays) in California and 14 days in New York.

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock-Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the CMA-A Composite Score, there was no statistically detectable difference between groups on the (a) fall assessment, but there were (b) statistically reliable differences between groups on means for the spring pre-kindergarten assessment (ES_S = .44, $p < .01$), and (c) in the rates of growth from fall pre-kindergarten to spring pre-kindergarten (difference in rates of growth = .08; ES_{Slope} = .3632, $p < .01$). Taken together, these three results provide clear evidence of an effect. There was (d) no difference in means for the spring kindergarten assessment, but there was (e) a statistically reliable difference in the rate of growth from spring pre-kindergarten to spring kindergarten (difference in rates of growth = -.04; ES_{Slope} = -.1690, $p < .01$). These last two results indicate that from spring of pre-kindergarten through spring of kindergarten, children who had been in *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* were learning at a *slower* rate relative to children who had been in the pre-kindergarten control classrooms. By the spring kindergarten assessment, there was no statistically detectable difference between the two groups. On the ANCOVA, there was a statistically reliable difference between groups in the covariate-adjusted spring pre-kindergarten means (ES_S = .35, $p < .01$), and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were statistically reliable differences between groups on: (a) the means from the fall assessment (ES_S = .25, $p < .05$), (b) the means from spring pre-kindergarten assessment (ES_S = .96, $p < .001$), and (c) the rates of growth from fall pre-kindergarten to spring pre-kindergarten (difference in rates of growth = .61; ES_{Slope} = .6999, $p < .0001$). The difference between groups at the fall assessment could reflect a failure of randomization to produce equivalent groups or an early treatment effect. We extrapolated back to the start of the school year; there was no statistically significant difference in means at the start of the school year. This finding suggests that the groups were equivalent at the beginning of the year and the observed difference in the fall reflected an early treatment effect. Taken together, these results provide evidence of a positive effect of *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* on Shape Composition relative to the control condition.

There was a statistically reliable difference between groups on (d) the means at the spring kindergarten assessment (ES_S = $p < .001$), and (e) the rates of growth from spring pre-kindergarten to spring kindergarten (difference in rates of growth = -.26; ES_{Slope} = -.2986, $p < .0001$). These last two results indicate that from spring of pre-kindergarten through spring of kindergarten, children who had been in the *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* were learning at a *slower* rate relative to children who had been in the pre-kindergarten control classrooms. Despite this slower rate of growth, the advantage obtained by the *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* children by the spring of pre-kindergarten remained through spring of the kindergarten year. On the ANCOVA, there were statistically reliable differences between groups in the covariate-adjusted spring pre-kindergarten means (ES_S = .91, $p < .0001$), and in the covariate-adjusted spring kindergarten means (ES_S = .30, $p < .01$).

Based the analyses for the three mathematics measures, we conclude that *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* had a positive effect on mathematics at the end of pre-kindergarten relative to the control condition.

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each reading assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the TERA, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the three reading measures, we conclude that *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* did not have a statistically detectable effect on reading relative to the control condition.

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted on the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Testing (CTOPP), Kindergarten, Elision subtest data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

For the ANCOVA on the kindergarten CTOPP data, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on these analyses for the phonological awareness measures, we conclude that *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatic Understanding subtest) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the TOLD Grammatic Understanding subtest, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the two language measures, we conclude that *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* did not have a statistically detectable effect on language development relative to the control condition.

Behavioral outcomes
Pre-kindergarten data from the three social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, we conducted an ANCOVA on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences between groups on the (a) fall pre-kindergarten assessment, or (b) spring pre-kindergarten assessment, but there was a statistically reliable difference between groups in the rate of growth from fall pre-kindergarten to spring pre-kindergarten (difference in rates of growth = 4.38, ES_{Slope} = .3040, $p < .05$). On the SSRS Social Skills scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the SSRS Problem Behaviors measure, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Problem Behaviors scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten.

On the PLBS, there were no statistically detectable differences in covariate-adjusted means on the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth between groups from fall to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten assessment.

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

Based on the analyses of the behavioral measures, we conclude that *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* did not have a statistically detectable effect on social and learning behaviors relative to the control condition.

Classroom Outcomes

The classroom-level effect sizes (ES_C) and slope effect sizes (ES_{Slope}) are presented in table A-18.

Overall classroom environment
We obtained observations on the Early Childhood Environment Rating Scale-Revised (ECERS-R) in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R, there were no statistically detectable differences between groups on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically significant difference was obtained on the spring pre-kindergarten observation.

Based on the analyses of the ECERS-R, we conclude that *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* did not have a statistically detectable effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, for each of the teacher-child relationship measures, ANCOVAs were conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the Arnett Detachment scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Harshness scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Permissiveness scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Positive Interactions scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses for the teacher-child relationship measures, we conclude that *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

Because the classroom instruction measures (Teacher Behavior Rating Scale [TBRS] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in spring pre-kindergarten, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software*

The impact of *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software* on the child- and classroom-level measures is summarized in table A-18.

Table A-18. Secondary analysis results for *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software*

Measure	RM analysis start of treatment[1]	RM analysis Fall Pre-K	RM analysis Spring Pre-K	Fall-Spring slope	ANCOVA[2] Spring Pre-K	RM analysis kindergarten	Spring Pre-K-Spring K slope	ANCOVA kindergarten
				Student-level effect sizes				
Mathematics								
WJ Applied Problems	.19	.19	.22	.0229	.16	.13	-.0437	.05
CMA-A Mathematics Composite	-.02	.07	.44**	.3632**	.35**	.13	-.1690**	.10
Shape Composition[3]	.08	.25*	.96***	.6999****	.91****	.41***	-.2986****	.30**
Reading								
TERA	.14	.14	.13	-.0099	.00	.31	.0969	.08
WJ Letter Word Identification	-.18	-.15	-.01	.1349	.06	.22	.1233	.21
WJ Spelling	.14	.15	.20	.0445	.17	.03	-.0881	-.08
Phonological awareness								
Pre-CTOPPP/CTOPP	-.17	-.13	.04	.1663	.11	†	†	-.11
Language								
PPVT	.02	.05	.17	.1223	.18	.11	-.0363	.09
TOLD	.12	.13	.17	.0385	.07	.08	-.0473	-.03
Behavior								
SSRS Social Skills	-.17	-.10	.22	.3040*	.24	†	†	.06
SSRS Problem Behavior[4]	.10	.07	-.09	-.1523	-.11	†	†	-.01
PLBS/LBS	-.10	-.06	.09	.1494	.08	†	†	.01
				Classroom-level effect sizes				
Global classroom quality								
ECERS-R	-.58	-.46	.05	.5040	-.22	†	†	†
Teacher-child interaction								
Arnett Detachment[5]	.12	.02	-.37	-.3841	.09	†	†	†
Arnett Harshness[5]	.16	.16	.18	.0174	.31	†	†	†
Arnett Permissiveness[5]	.12	.02	-.45	-.4521	-.36	†	†	†
Arnett Positive Interactions	-.27	-.19	.16	.3471	-.24	†	†	†

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.
* $p < .05$; ** $p < .01$; *** $p < .001$; **** $p < .0001$
[1] The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).
[2] The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.
[3] Building Blocks, Shape Composition task
[4] Higher scores on this scale represent more negative child behaviors.
[5] Lower scores on this scale represent a more positive classroom environment.
NOTE: RM: Repeated Measures
ANCOVA: Analysis of covariance
Refer to the glossary for abbreviations of the measures.
SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Project Approach: Purdue University and University of Wisconsin-Milwaukee (Wisconsin site)

We present analyses for each of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. Within each domain, we present the repeated measures models followed by the ANCOVAs that included the fall assessment as one of the covariates in the model. The student-level effect sizes (ES_S) and slope effect sizes (ES_{Slope}) are presented in table A-19.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 13 days (including Saturdays, Sundays, and holidays).

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock-Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, however, there was a statistically significant difference in the (e) rate of growth from spring pre-kindergarten to spring kindergarten (difference in rate of growth = 2.37; ES_{Slope} = .1043, $p < .05$). On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the CMA-A Composite Score, there were no statistically detectable differences between groups on the (a) fall assessment or (b) spring pre-kindergarten assessment; however, there was a statistically reliable difference in the (c) rates of growth from fall pre-kindergarten to spring pre-kindergarten (difference in rate of growth = .06; ES_{Slope} = .2640, $p < .05$). In this instance, we do not have all three conditions necessary to indicate statistical evidence of a treatment effect on the WJ Applied Problems. There were no statistically detectable differences between groups on (d) the spring kindergarten assessment or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no detectable significant differences between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment or the (c) rates of growth from fall pre-kindergarten to spring pre-kindergarten. There were no statistically detectable differences between groups on the (d) spring kindergarten assessment or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

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

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each reading assessment, an ANCOVA was conducted in which the covariates were: fall assessment score, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the TERA, there were (a) no statistically detectable differences between groups on the (a) fall assessment or (b) spring pre-kindergarten assessment; however, there was a statistically reliable difference between groups on the (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten (difference in rate of growth = -2.1; ES_{Slope} = -.2202, $p < .05$). We do not have all three conditions necessary to indicate statistical evidence of a treatment effect on the TERA. There was no statistically detectable difference between groups on (d) the spring kindergarten assessment or the (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there were no statistically significant differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, and (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. There was no statistically significant difference on the (d) spring kindergarten assessment; however, there was a statistically significant difference between groups in the (e) rate of growth from spring pre-kindergarten to spring kindergarten (difference in rate of growth = -5.7; ES_{Slope} = -.2102, $p < .01$). On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, but there was a statistically significant difference between groups in the covariate-adjusted spring kindergarten means (ES_S = -.44, $p < .05$).

On the WJ Spelling test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

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

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted on the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there was no statistically detectable difference between groups in the covariate-adjusted spring pre-kindergarten means. For the ANCOVA on the kindergarten CTOPP data, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the phonological awareness measures, we conclude that *Project Approach* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT] and Test of Language Development [TOLD] Grammatic Understanding subtest) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there were no statistically detectable differences between groups on the (a) fall assessment and (b) the spring pre-kindergarten assessment, or (c) the rate of growth from fall pre-kindergarten to spring pre-kindergarten. There were no statistically detectable differences between groups on (d) the spring kindergarten assessment or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the TOLD Grammatic Understanding subtest, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten, the spring kindergarten assessment, or the (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten.

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

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (Social System Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, we conducted an ANCOVA on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Social Skills scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten, but there was a significant difference between groups at spring kindergarten ($ES_S = -.44$, $p < .05$).

On the SSRS Problem Behaviors measure, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Problem Behaviors scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for the spring pre-kindergarten assessment, but there was a significant difference between groups at spring kindergarten ($ES_S = .49$, $p < .05$).

On the PLBS, there were no statistically detectable differences in covariate-adjusted means on the (a) fall pre-kindergarten assessment or (b) spring pre-kindergarten assessment. There was a statistically significant difference in the rate of growth from fall to spring pre-kindergarten (different in rates of growth = -2.9, $ES_{Slope} = -.2922$, $p < .05$). On the ANCOVA, there was a statistically detectable difference in the covariate-adjusted means for the spring pre-kindergarten assessment ($ES_S = -.37$, $p < .05$).

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

Based on the analyses of the behavioral measures, we conclude that *Project Approach* did not have a statistically detectable effect on social and learning behaviors relative to the control condition during pre-kindergarten, but *Project Approach* had a negative effect on behavior by spring of the kindergarten year relative to the control condition.

Classroom Outcomes

The classroom-level effect sizes (ES_C) and slope effect sizes (ES_{Slope}) are presented in table A-19.

Overall classroom environment

We obtained observations on the Early Childhood Environment Rating Scale-Revised (ECERS-R) in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses of the ECERS-R, we conclude that *Project Approach* did not have a statistically detectable on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, for each of the teacher-child relationship measures, ANCOVAs were conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the Arnett Detachment scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Harshness scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Permissiveness scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, there was no statistically detectable difference on the spring pre-kindergarten observation.

On the Arnett Positive Interactions scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses of the four teacher-child relationship measures, we conclude that *Project Approach* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

Because the classroom instruction measures (Teacher Behavior Rating Scale [TBRS] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in spring pre-kindergarten, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for *Project Approach*

The impact of *Project Approach* on the child- and classroom-level measures is summarized in table A-19.

Table A-19. Secondary analysis results for *Project Approach*

Measure	RM analysis start of treatment[1]	RM analysis Fall Pre-K	RM analysis Spring Pre-K	Fall-Spring slope	ANCOVA[2] Spring Pre-K	RM analysis kindergarten	Spring Pre-K-Spring K slope	ANCOVA kindergarten
Student-level effect sizes								
Mathematics								
WJ Applied Problems	-.08	-.05	.07	.1201	-.11	.27	.1043*	.02
CMA-A Mathematics Composite	-.15	-.09	.18	.2640*	.00	.22	.0240	-.03
Shape Composition[3]	-.04	.02	.27	.2478	-.13	.24	-.0173	-.11
Reading								
TERA	.42	.36	.14	-.2202*	-.23	.29	.0807	-.18
WJ Letter Word Identification	.49	.47	.42	-.0548	-.05	.03	-.2102**	-.44*
WJ Spelling	.34	.33	.27	-.0561	-.19	.14	-.0682	-.35
Phonological awareness								
Pre-CTOPPP/CTOPP	.15	.13	.05	-.0752	-.27	†	†	-.17
Language								
PPVT	-.06	-.02	.16	.1775	.07	.10	-.0352	-.10
TOLD	-.07	-.03	.15	.1752	-.08	.32	.0875	.04
Behavior								
SSRS Social Skills	-.07	-.05	.04	.0891	.06	†	†	-.44*
SSRS Problem Behavior[4]	.34	.37	.50	.1212	.23	†	†	.49*
PLBS/LBS	.06	-.01	-.31	-.2922*	-.37*	†	†	-.42
Classroom-level effect sizes								
Global classroom quality								
ECERS-R	1.07	.84	-.19	-1.0030	-.27	†	†	†
Teacher-child interaction								
Arnett Detachment[5]	.12	.20	.57	.3596	.37	†	†	†
Arnett Harshness[5]	1.10	1.06	.86	-.1903	-.21	†	†	†
Arnett Permissiveness[5]	-.28	-.30	-.43	-.1191	-.57	†	†	†
Arnett Positive Interactions	-.98	-.98	-.99	-.0091	.69	†	†	†

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

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

[1] The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).

[2] The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.

[3] Building Blocks, Shape Composition task

[4] Higher scores on this scale represent more negative child behaviors.

[5] Lower scores on this scale represent a more positive classroom environment.

NOTE: RM: Repeated Measures
ANCOVA: Analysis of covariance

Refer to the glossary for abbreviations of the measures.

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

Project Construct: University of Missouri-Columbia (Missouri site)

We present analyses for each of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. Within each domain, we present the repeated measures models followed by the ANCOVAs that included the fall assessment as one of the covariates in the model. The student-level effect sizes (ES_S) and slope effect sizes (ES_{Slope}) are presented in table A-20.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 42 days (including Saturdays, Sundays, and holidays).

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock-Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the CMA-A Composite Score, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there were no statistically detectable differences between groups on (a) the fall assessment, or (c) in the rate of growth from fall pre-kindergarten to spring pre-kindergarten, but there was (b) a significant difference between groups on the spring pre-kindergarten assessment (ES_S = -.42, p < .01). In this instance, we do not have all three conditions necessary to indicate statistical evidence of a treatment effect on the Shape Composition task. There was no statistically detectable difference on the spring kindergarten assessment, but there was a statistically significant difference between groups in the rate of growth from spring pre-kindergarten to spring kindergarten (difference in rate of growth = .01; ES_S = .2846, p < .001). The observed difference in rate of growth from spring pre-kindergarten to spring kindergarten with no statistically significant difference in means on the spring pre-kindergarten and spring kindergarten assessments provides inconclusive evidence that children from *Project Construct* were learning at a slower rate in the year after the pre-kindergarten intervention. On the ANCOVA, there was a statistically significant difference between groups in the covariate-adjusted spring pre-kindergarten means (ES_S = -.44, p < .05), and no statistically significant differences between groups in the covariate-adjusted spring kindergarten means.

The ANCOVA analysis indicates that from the fall assessment to the spring pre-kindergarten assessment students did not gain as much relative to students in the control classrooms.

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

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each reading assessment, an ANCOVA was conducted in which the covariates were: fall assessment score, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the TERA, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

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

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted on the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

For the ANCOVA on the kindergarten CTOPP data, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses of the phonological awareness measures, we conclude that *Project Construct* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT], and TOLD Grammatic Understanding subtest) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the TOLD Grammatic Understanding subtest, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten.

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

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, ANCOVA analyses were conducted on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no statistically detectable differences between groups on the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Social Skills scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the SSRS Problem Behaviors measure, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Problem Behaviors scale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the PLBS, there were no statistically detectable differences in covariate-adjusted means on the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth between groups from fall to spring pre-kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring pre-kindergarten assessment.

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

Based on the analyses of the behavioral measures, we conclude that *Project Construct* did not have a statistically detectable effect on social and learning behaviors relative to the control condition.

Classroom Outcomes

The classroom-level effect sizes (ES_C) and slope effect sizes (ES_{Slope}) are presented in table A-20.

Overall classroom environment

We obtained observations on the Early Childhood Environment Rating Scale-Revised (ECERS-R) in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with the following covariates: teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

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

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, for each of the teacher-child relationship measures, ANCOVAs were conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the Arnett Detachment scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Harshness scale, there were no statistically detectable differences on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Permissiveness scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Positive Interactions scale, there were no statistically detectable differences in the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses for the four teacher-child relationship measures, we conclude that *Project Construct* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

Because the classroom instruction measures (Teacher Behavior Rating Scale [TBRS] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained in spring pre-kindergarten, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for *Project Construct*

The impact of *Project Construct* on the child- and classroom-level measures is summarized in table A-20.

Table A-20. Secondary analysis results for *Project Construct*

Measure	RM analysis start of treatment[1]	RM analysis Fall Pre-K	RM analysis Spring Pre-K	Fall-Spring slope	ANCOVA[2] Spring Pre-K	RM analysis kindergarten	Spring Pre-K-Spring K slope	ANCOVA kindergarten
				Student-level effect sizes				
Mathematics								
WJ Applied Problems	-.07	-.04	.06	.0990	.10	.08	.0107	.07
CMA-A Mathematics Composite	-.02	-.04	-.11	-.0725	-.07	-.06	.0287	-.08
Shape Composition[3]	-.10	-.16	-.42**	-.2492	-.41*	.12	.2846***	.13
Reading								
TERA	-.07	-.06	.00	.0576	.18	-.03	-.0166	.12
WJ Letter Word Identification	.00	-.01	-.05	-.0370	-.07	.16	.1129	.22
WJ Spelling	-.31	-.28	-.15	.1268	-.02	.00	.0798	.13
Phonological awareness								
Pre-CTOPPP/CTOPP	.37	.32	.10	-.2176	-.07	†	†	-.12
Language								
PPVT	.02	.02	.03	.0129	-.03	.10	.0339	.11
TOLD	.23	.18	-.05	-.2196	-.13	.01	.0305	-.08
Behavior								
SSRS Social Skills	.22	.22	.22	.0005	.07	†	†	.12
SSRS Problem Behavior[4]	-.22	-.19	-.08	.1060	.07	†	†	.07
PLBS/LBS	.20	.16	.00	-.1585	-.11	†	†	-.02
				Classroom-level effect sizes				
Global classroom quality								
ECERS-R	.27	.32	.54	.2112	.35	†	†	†
Teacher-child interaction								
Arnett Detachment[5]	.27	.24	.12	-.1183	-.24	†	†	†
Arnett Harshness[5]	-.95	-.80	-.13	.6485	-.02	†	†	†
Arnett Permissiveness[5]	.42	.34	-.02	-.3532	-.18	†	†	†
Arnett Positive Interactions	.40	.41	.46	.0464	.39	†	†	†

† Not applicable. Four of the kindergarten student-level measures were not on the same scale as the pre-kindergarten measures. The classroom-level data were only collected during the pre-kindergarten year of the study.
* $p < .05$; ** $p < .01$; *** $p < .001$
[1] The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).
[2] The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.
[3] Building Blocks, Shape Composition task
[4] Higher scores on this scale represent more negative child behaviors.
[5] Lower scores on this scale represent a more positive classroom environment.
NOTE: RM: Repeated Measures
ANCOVA: Analysis of covariance
Refer to the glossary for abbreviations of the measures.
SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Ready, Set, Leap!: University of California, Berkeley (New Jersey site)

We present analyses for each of the child-level measures (i.e., the mathematics, reading, phonological awareness, and language assessments) followed by the analyses of the classroom observation data. Within each domain, we present the repeated measures models followed by the ANCOVAs that included the fall assessment as one of the covariates in the model. The student-level effect sizes (ES_S) and slope effect sizes (ES_{Slope}) are presented in table A-21.

To provide contextual information for judging the possibility of early treatment effects, the lag between the start of treatment to the beginning of the child assessment window was 35 days (including Saturdays, Sundays, and holidays).

Child Outcomes

Mathematics assessments

We used repeated measures linear spline models to analyze the data from all three mathematics measures (Woodcock-Johnson [WJ] Applied Problems, Child Math Assessment-Abbreviated [CMA-A] Composite Score, and Shape Composition). For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each mathematics assessment, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the WJ Applied Problems, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the CMA-A Composite Score, there were (a) no statistically detectable differences between groups on the fall assessment, (b) a statistically reliable difference on the spring pre-kindergarten assessment (ES_S = -.24, $p < .05$), and (c) no statistically detectable difference in rate of growth from fall pre-kindergarten to spring pre-kindergarten. Taken together, these three results do not provide conclusive evidence of an effect of *Ready, Set, Leap!* relative to the control condition on the CMA-A Composite Score for the pre-kindergarten year. In addition, there were no statistically detectable differences between groups on (d) the spring kindergarten assessment or (e) the rates of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the Shape Composition task, there was (a) a statistically reliable difference between groups on the fall assessment (ES_S = .25, $p < .05$), but none on the (b) spring pre-kindergarten assessment or (c) the rates of growth from fall pre-kindergarten to spring pre-kindergarten. The difference between groups on the fall assessment could reflect either the failure of randomization to produce equivalent groups or an early treatment effect (which was not sustained to the spring pre-kindergarten assessment). To examine the possibility of an early treatment effect, we extrapolated back to the beginning of the school year and found a statistically reliable difference favoring the *Ready, Set, Leap!* group on the Shape Composition measure (ES_S = .29, $p < .05$). There was no statistically significant difference on the start of treatment extrapolated means on any of the other mathematics measures or other child measures. Given the lack of consistent results across measures and because the test of the extrapolated means does not provide conclusive evidence of nonequivalence at the start of the treatment,[10] we conclude the evidence is inconclusive for determining nonequivalence at the start of treatment.

In addition, on the Shape Composition task, there were no statistically significant differences between groups on (d) the spring kindergarten assessment or (e) the rates of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically significant differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically significant differences between groups in the covariate-adjusted spring kindergarten means.

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

Reading assessments

Data from the three reading measures (Test of Early Reading Ability [TERA], WJ Letter Word Identification, and WJ Spelling) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, for each reading assessment, an ANCOVA was conducted in which the covariates were: fall assessment score, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the TERA, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

For the WJ Letter Word Identification test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

On the WJ Spelling test, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means, and no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the three reading measures, we conclude that *Ready, Set, Leap!* did not have a statistically detectable effect on reading relative to the control condition.

---

[10] For the following reasons, this difference suggests, but does not prove, nonequivalence of treatment and control groups. The extrapolation is based on the average score for each group for the fall assessment and the rate of growth from fall to spring for each group. Although the slope from fall to spring is our best estimate of the rate of growth from the beginning of the school year to the fall assessment, if there were an early treatment effect, there is no theoretical reason to assume that growth was constant from the start of the school year to the spring assessment. Across the nine academic outcomes, the Shape Composition was the only one on which there were statistically reliable differences for the fall assessment means or the start of treatment extrapolated means.

Phonological awareness

We conducted a repeated measures analysis of pre-kindergarten data from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest. For this analysis, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: fall assessment score was not included). In addition, ANCOVA analyses were conducted on the pre-kindergarten Pre-CTOPPP data and the kindergarten Comprehensive Test of Phonological Processing (CTOPP), Kindergarten, Elision subtest data with the following covariates: *Pre-CTOPPP fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the Pre-CTOPPP, there were no statistically detectable differences between groups on the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the ANCOVA, there were no statistically detectable differences between groups in the covariate-adjusted spring pre-kindergarten means.

For the ANCOVA on the kindergarten CTOPP data, there were no statistically detectable differences between groups in the covariate-adjusted spring kindergarten means.

Based on the analyses for the phonological awareness measures, we conclude that *Ready, Set, Leap!* did not have a statistically detectable effect on phonological awareness relative to the control condition.

Language assessments

Data from the two language measures (Peabody Picture Vocabulary Test [PPVT], Test of Language Development [TOLD] Grammatic Understanding subtest) were analyzed using repeated measures linear spline models. For each model, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, an ANCOVA was conducted in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

For the PPVT, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the TOLD Grammatic Understanding subtest, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, (c) spring kindergarten assessment, (d) rate of growth from fall pre-kindergarten to spring pre-kindergarten, or (e) rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

Based on the analyses for the two language measures, we conclude that *Ready, Set, Leap!* did not have a statistically detectable effect on language development relative to the control condition.

Behavioral outcomes

Pre-kindergarten data from the three social behavioral measures (Social Skills Rating System [SSRS] Social Skills scale, SSRS Problem Behaviors scale, and Preschool Learning Behaviors Scale [PLBS]) were analyzed using simple repeated measures models. For each of these models, we included the following covariates: child age, gender, race/ethnicity, disability status as reported by parent, and mother's education (note: no fall assessment score was included). In addition, we conducted an ANCOVA on the pre-kindergarten (SSRS Social Skills scale, SSRS Problem Behaviors scale, and PLBS) and kindergarten (SSRS Social Skills, SSRS Problem Behaviors, and Learning Behaviors Scale [LBS]) data in which the covariates were: *fall assessment score*, child age, gender, race/ethnicity, disability status as reported by parent, and mother's education.

On the SSRS Social Skills measure, there were no differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Social Skills subscale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten.

On the SSRS Problem Behaviors measure, there were no statistically detectable differences between groups on the (a) fall assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth from fall pre-kindergarten to spring pre-kindergarten. On the SSRS Problem Behaviors subscale, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted and we could not test the rate of growth from spring pre-kindergarten to spring kindergarten. On the ANCOVA, there were no statistically detectable differences in the covariate-adjusted means for spring pre-kindergarten or spring kindergarten assessments.

On the PLBS, there were no statistically detectable differences in covariate-adjusted means on the (a) fall pre-kindergarten assessment, (b) spring pre-kindergarten assessment, or (c) rate of growth between groups from fall to spring pre-kindergarten. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring pre-kindergarten assessment.

On the LBS, because the measure changed from pre-kindergarten to kindergarten, a repeated measures analysis was not conducted. On the ANCOVA, there was no statistically detectable difference in the covariate-adjusted means for the spring kindergarten assessment.

Based on the analyses of the behavioral measures, we conclude that *Ready, Set, Leap!* did not have a statistically detectable effect on social and learning behaviors relative to the control condition.

Classroom Outcomes

The classroom-level effect sizes (ES_C) and slope effect sizes (ES_{Slope}) are presented in table A-21.

Overall classroom environment

We obtained observations on the Early Childhood Environment Rating Scale-Revised (ECERS-R) in the fall and spring of the pre-kindergarten year and conducted repeated measures analyses with the following covariates: teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site (note: no fall observation score was included). In addition, an ANCOVA was conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the ECERS-R, there were no statistically detectable differences between groups on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses of the ECERS-R, we conclude that *Ready, Set, Leap!* did not have a statistically detectable effect on overall classroom quality relative to the control condition.

Teacher-child relationships

We obtained observations on the Arnett Detachment, Harshness, Permissiveness, and Positive Interactions scales in fall and spring of the pre-kindergarten year and conducted repeated measures analyses with teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as covariates (note: no fall observation score was included). In addition, for each of the teacher-child relationship measures, ANCOVAs were conducted with the *fall observation score*, teacher has a BA degree, previous teaching experience, teacher's race/ethnicity, child/adult ratio in classroom, average class size, city size, and site as the covariates.

On the Arnett Detachment scale, there were no statistically detectable differences between groups on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Harshness scale, there were no statistically detectable differences between groups on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

On the Arnett Permissiveness scale, there were no statistically detectable differences between groups on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained for the spring pre-kindergarten observation.

On the Arnett Positive Interactions scale, there were no statistically detectable differences between groups on the (a) covariate-adjusted means for the fall pre-kindergarten observation, (b) spring pre-kindergarten observation, or (c) rate of change from the fall to spring observation. On the ANCOVA, no statistically detectable difference was obtained on the spring pre-kindergarten observation.

Based on the analyses for the four teacher-child relationship measures, we conclude that *Ready, Set, Leap!* did not have a statistically detectable effect on teacher-child relationships relative to the control condition.

Classroom instruction

Because the classroom instruction measures (Teacher Behavior Rating Scale [TBRS] Book Reading, Print and Letter Knowledge, Written Expression, Phonological Awareness, Oral Language, and Math Concepts) were only obtained during the spring of pre-kindergarten, neither the repeated measures nor an ANCOVA including a fall observation as a covariate was conducted. Hence, no additional analyses beyond what was reported in the body of the report were conducted.

Summary of Results for *Ready, Set, Leap!*

The impact of *Ready, Set, Leap!* on the child- and classroom-level measures is summarized in table A-21.

Table A-21. Secondary analysis results for *Ready, Set, Leap!*

Measure	RM analysis start of treatment[1]	RM analysis Fall Pre-K	RM analysis Spring Pre-K	Fall-Spring slope	ANCOVA[2] Spring Pre-K	RM analysis kindergarten	Spring Pre-K-Spring K slope	ANCOVA kindergarten
Student-level effect sizes								
Mathematics								
WJ Applied Problems	.14	.12	.04	-.0753	.03	.00	-.0221	-.05
CMA-A Mathematics Composite	-.07	-.10	-.24*	-.1317	-.23	-.10	.0719	-.11
Shape Composition[3]	.29*	.25*	.08	-.1665	-.01	.03	-.0277	-.06
Reading								
TERA	.06	.07	.08	.0125	.07	.01	-.0361	-.05
WJ Letter Word Identification	-.03	-.02	.01	.0360	.03	-.12	-.0728	-.06
WJ Spelling	.01	.04	.20	.1477	.13	.04	-.0829	.02
Phonological awareness								
Pre-CTOPPP/CTOPP	.03	.01	-.09	-.0913	-.06	†	†	-.02
Language								
PPVT	.19	.18	.15	-.0320	.01	-.02	-.0888	-.13
TOLD	.07	.04	-.11	-.1425	-.14	-.03	.0441	-.04
Behavior								
SSRS Social Skills	-.01	-.02	-.05	-.0317	-.05	†	†	-.03
SSRS Problem Behavior[4]	.02	.01	-.03	-.0373	.00	†	†	.07
PLBS/LBS	.10	.09	.07	-.0279	.01	†	†	-.01
Classroom-level effect sizes								
Global classroom quality								
ECERS-R	-.03	.01	.16	.1473	.32	†	†	†
Teacher-child interaction								
Arnett Detachment[5]	.19	.19	.19	.0028	.17	†	†	†
Arnett Harshness[5]	.23	.25	.30	.0561	.26	†	†	†
Arnett Permissiveness[5]	-.36	-.34	-.24	.0903	-.09	†	†	†
Arnett Positive Interactions	-.08	-.05	.04	.0936	.15	†	†	†

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

* $p < .05$

[1] The values represent the extrapolated scores back to the beginning of the school year (i.e., start of treatment).

[2] The reported effect sizes from the ANCOVA analyses may be biased downward because of early treatment effects.

[3] Building Blocks, Shape Composition task

[4] Higher scores on this scale represent more negative child behaviors.

[5] Lower scores on this scale represent a more positive classroom environment.

NOTE: RM: Repeated Measures
ANCOVA: Analysis of covariance

Refer to the glossary for abbreviations of the measures.

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

Appendix B:
Data Analysis Approach and Statistical Model

Research Questions

The Preschool Curriculum Evaluation Research (PCER) initiative focused on the impact of the intervention curricula on students' reading, phonological awareness, early language, early mathematics knowledge, and behavior (including social skills) at the end of pre-kindergarten and kindergarten. These domains of knowledge and skills are predictive of academic success in the early years of elementary school (Downer and Pianta 2006; Miles and Stipek 2006).

In addition, the PCER evaluation study also examined the impact of the curriculum interventions on teachers' classroom instructional practice, teacher-child interaction, and global classroom quality. These dimensions of early childhood programs have been posited as mediators (e.g., instructional practice) and moderators (e.g., teacher-child interaction, classroom quality) of the relation between early childhood curricula and child outcomes (Arnett 1989; Cost, Quality, and Child Outcomes Study 1995; Peisner-Feinberg and Burchinal 1997; Ruopp et al. 1979).

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

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

Data Collection, Sample, and Assignment

Data were collected in the fall (baseline) and spring of the pre-kindergarten year and in the spring of the kindergarten year to answer the research questions outlined above. Each research team recruited preschool programs, teachers, children, and parents for participation in the PCER evaluation study. Overall, 2,911 children, 315 preschool classrooms, and 208 preschools were involved in the PCER initiative.

The research teams recruited samples of convenience from local preschool programs willing to agree to the random assignment of classrooms or schools to treatment and control conditions. Table B-1 details the research teams, the curricula they evaluated, and the sample size and unit of assignment they used.

Variables Used in Analysis and General Data Issues

Time Structure of Data

Data were collected at three time points: fall pre-kindergarten (i.e., baseline), spring pre-kindergarten, and spring kindergarten. A common measurement battery was used to collect child, parent, teacher, and classroom observation (pre-kindergarten only) data at all research sites. For a few child-outcome measures, the assessment instruments changed between pre-kindergarten and kindergarten or changes in standardizations. Instruments included the Social Skills Rating System (SSRS) Socials and Problem Behaviors scales; the Preschool Learning Behaviors Scale (PLBS) (pre-kindergarten); the Learning Behaviors Scale (LBS) (kindergarten); the Preschool Comprehensive Test of Phonological and Print Processing, Elision subtest (Pre-CTOPPP) (pre-kindergarten); and the Comprehensive Test of Phonological Processing (CTOPP), Elision subtest. Classrooms were only observed during the intervention (pre-kindergarten) year, as the children all

Table B-1. Units of random assignment for evaluation of each curriculum

Research team (site)	Curricula	Treatment sample	Control sample	Students
Vanderbilt University (TN)	*Bright Beginnings*	7 classrooms	7 classrooms	T: 103 C: 105
	Creative Curriculum	7 classrooms		T: 101
UNC-Charlotte (NC, GA)	*Creative Curriculum*	9 classrooms	9 classrooms	T: 97 C: 97
University of New Hampshire (NH)	*Creative Curriculum with Ladders to Literacy*	7 classrooms	7 classrooms	T: 62 C: 61
Success for All Foundation (NJ, KS, FL)	*Curiosity Corner*	10 Pre-K programs	8 Pre-K programs	T: 105 C: 110
University of Texas Health Science Center at Houston (TX)	*Doors to Discovery*	14 classrooms	15 classrooms	T: 101 C: 96
	Let's Begin with the Letter People	15 classrooms		T: 100
University of North Florida (FL)	*Early Literacy and Learning Model*	14 classrooms[1]	14 classrooms[1]	T: 137 C: 107
University of Virginia (VA)	*Language-Focused Curriculum*	7 classrooms	7 classrooms	T: 97 C: 98
Florida State University (FL)	*DLM Early Childhood Express with Open Court Reading Pre-K*	5 Pre-K programs	6 Pre-K programs	T: 101 C: 97
	Literacy Express	6 Pre-K programs		T: 99
UC-Berkeley and University at Buffalo, SUNY (CA, NY)	*Pre-K Mathematics with DLM Early Childhood Express Math software*	20 classrooms	20 classrooms	T: 159 C: 157
Purdue University and University of WI-Milwaukee (WI)	*Project Approach*	7 classrooms	6 classrooms	T: 114 C: 90
University of Missouri-Columbia (MO)	*Project Construct*	10 Pre-K programs[1]	11 Pre-K programs[1]	T: 123 C: 108
UC-Berkeley (NJ)	*Ready, Set, Leap!*	18 classrooms	21 classrooms	T: 149 C: 137

[1] After one program or classroom attrited.

NOTE: T: Treatment Group
C: Control Group

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

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

dispersed to a large number of kindergarten classrooms in the second year of the study. Classroom observation data were collected in the fall and spring of the pre-kindergarten year using the Early Childhood Environment Rating Scale-Revised (ECERS-R) and the Arnett Caregiver Interaction Scale (Arnett) measures. The Teacher Behavior Rating Scale (TBRS) was added to the classroom observation battery for the pre-kindergarten spring observation time point.

Outcome Variables

A total of 12 child-outcome measures and 11 intermediate outcome measures were used in the data analyses to answer the research questions outlined above. The measures included in the impact analyses are listed in table B-2. For the Woodcock Johnson (WJ) measures (Letter-Word Identification, Applied Problems, and Spelling subscales), the "W" scores were used in the analyses because the WJ standardized scores account for developmental growth associated with the child's age in years. "W" scores are a special transformation of the Rasch ability scale. The "W" score for each test is centered on a value of 500, which has been set to the approximate performance of a 10-year-old student. Individuals whose performance on the measure is lower than the average score receive scores below 500. The use of the Rasch score allows researchers to record changes in actual ability within or across years of a study. For the Peabody Picture Vocabulary Test (PPVT), the Test of Early Reading Ability (TERA), the Test of Language Development (TOLD) Grammatic Understanding subtest, and the Pre-CTOPPP/CTOPP, Elision subtest measures, raw scores were used in the analyses. Standard scores were used for the SSRS measures, and standard T-scores were used for PLBS/LBS measures. T-scores are test scores converted to an equivalent standard score in a normal distribution with a mean of 50 and a standard deviation of 10.

Covariates

In all models used in the Main and Secondary analyses, covariates were included to increase the precision of the impact estimates by adjusting for chance baseline differences between the treatment and control groups on the characteristics represented by the covariates. In our models of child-outcome measures, the following set of covariates was used:

- child's age;
- child's race/ethnicity;
- child's gender;
- self-reported maternal education;
- proxy disability status indicator (parent-reported IEP);
- site within research team (if a team had more than one site); and
- curriculum within research team (if a team evaluated more than one curriculum).

The following additional covariates were included in the repeated measures models:

- intervention exposure ($Time_1$) (i.e., time between start of intervention and the child's assessment, in *all repeated measures models)*, and
- post-intervention time($Time_2$) (i.e., time since the spring assessment of the pre-kindergarten year in *repeated measures spline model only)*.

The following additional covariate was included in the ANCOVA models:

- child's baseline score on the relevant child-outcome measure.

Table B-2. Variables in analysis

Child outcome measures	Classroom outcome measures
• 3 mathematics measures: o WJ Applied Problems o CMA-A Mathematics Composite o Shape Composition[1] • 1 phonological awareness measure: o Pre-CTOPPP/CTOPP • 3 reading measures: o WJ Letter Word Identification o WJ Spelling o TERA • 2 language measures: o TOLD o PPVT • 3 behavior measures: o SSRS Social Skills o SSRS Problem Behaviors o PLBS/LBS	• 3 classroom observation measures: o ECERS-R o Arnett Subscales ▪ Detachment ▪ Harshness ▪ Permissive ▪ Positive Interaction o TBRS: (Spring pre-kindergarten only) ▪ Book Reading ▪ Oral Language use ▪ Print and Letter Knowledge ▪ Written Expression ▪ Phonological Awareness ▪ Math Concepts

[1] Building Blocks, Shape Composition task
NOTE: Refer to the glossary for abbreviations of the measures.
SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

For the classroom models the following covariates were included:

- teacher has a BA degree;
- previous teaching experience;
- teacher's race/ethnicity;
- child/adult ratio in classroom;
- average number of students attending classroom;
- city size;
- site within research team (if a team had more than one site); and
- curriculum within research team (if a team evaluated more than one curriculum).

The following additional covariate was included in the repeated measures models:

- intervention exposure ($Time_1$) (i.e., time between start of intervention and the classroom assessment, in all repeated measures models).

The following additional covariate was included in the ANCOVA models:

- baseline scores on the ECERS-R and Arnett measures were included in the models for these classroom outcomes.

Covariates Excluded From the Child-Level Analysis Models

Preliminary analyses were conducted to determine which covariates should be included in the child-outcome measures models. The analysis results indicated that the following list of covariates were unrelated to the child measures under consideration and accounted for little or no portion of the explained variance:

- child born in the United States;
- mother works full-time;
- two parents in the household;
- parents' income level;
- parents' age;
- primary language used in the home;
- mother reads daily;
- parents' health;
- mother depressed;
- child read to daily;
- number of types of children's reading material in home;
- child watches TV more than two hours per day;
- number of rules;
- number of weekly activities;
- number of monthly activities;
- teacher has AA degree;
- teacher has specialization in preschool;
- teacher has Child Development Associate (CDA) credential;
- classroom teaching experience;
- preschool teaching experience;
- teacher's salary level;
- teacher's age;
- teacher developmental attitudes;
- teacher didactic activities;
- child/adult ratio in classroom;
- number of child's absences;
- teacher reported extent of behavior problems in classroom;
- ECER-R total score; and
- Arnett total score.

Covariates Excluded from the Classroom-Level Analysis Models

Preliminary classroom level analyses looked at and eliminated the covariates listed below. As with the child covariates, these covariates were shown to be unrelated to classroom measures and accounted for little or no portion of the explained variance. The following covariates were not included in the final classroom models:

- whether or not the teacher has an AA degree;
- what specialization the teacher has;
- whether the teacher is credentialed in child development;
- classroom teaching experience;
- teacher's salary level;
- developmental activities (composite from Teacher Interview);
- didactic activities (composite from Teacher Interview); and
- teacher reported extent of behavior problems of classroom.

Intervention/Control Group Assignment and Coding

An "intention-to-treat" logic[1] was employed for the site-level/curriculum-specific analyses of program impacts for child and classrooms measures. All children at a given site were included in the analysis in the group to which they were randomized, even if they were lost to follow-up for some reason.

To accurately compute differences between control and intervention groups, we utilized an intervention-within-sites effect-coding scheme. That is, for grantees with one site and one intervention, the intervention group was coded as +.05 and the control group was coded as -.05; for grantees with more than one site, dummy variables were created for each site that were coded +.05 and -.05 for intervention and control groups as above and coded 0 for information collected at the other site(s). This approach ensures that for sites that implemented only one curriculum, the significance of the intervention versus control difference is reflected in the parameter representing a particular intervention-within-site and that the control groups within sites were used as the comparison for their respective interventions. In cases where a grantee examined more than one intervention (i.e., Vanderbilt University, University of Texas Health Science Center at Houston [University of Texas-Houston] and Florida State University [FSU]) the effects of the shared control group were taken into account when reporting Intervention versus control differences. To obtain accurate estimates of the intervention effect, contrasts and estimates in SAS (Statistical Analysis Software) were used to provide the specified coefficients of interest.

Missingness—Attrition and Response Rates

Some attrition occurred in each of the 2 years in which data were collected, pre-kindergarten and kindergarten. The attrition rates were generally low and there was no evidence of differential attrition (i.e., attrition rates did not differ significantly between intervention and control groups). Response rates were lowest for the parent interview data collected by some research teams, and for the teacher reports at the end of the kindergarten year (see table B-3). The software used, SAS PROC MIXED, is designed to utilize all available data, even if some of the observations at a given time point are missing. In the two longitudinal models, if a child's data were missing at one of the two follow-up data collection time points, the child's earlier collected data were still included and used in estimation, where appropriate.

[1] "Intention to treat" is a strategy for the analysis of randomized controlled trials that compares individuals in the groups to which they were originally randomly assigned. It is based on the assumption that all of the individuals assigned to the intervention group may not receive the full exposure to the treatment, even though this was the original intent of the intervention.

Table B-3. Response rates and attrition

Research team	Response rate Fall 2003	Percent of sample with data Spring 2004	Percent of sample with data Spring 2005
Vanderbilt (n = 309)			
Child Assessments	100	94	97
Teacher Report	100	90	90
Parents Interview	82	81	75
UNC-Charlotte (n = 194)			
Child Assessments	98	88	85
Teacher Report	100	88	56
Parents Interview	87	69	71
University of New Hampshire (n = 123)			
Child Assessments	100	85	66
Teacher Report	99	81	50
Parents Interview	16	45	51
Success for All (n = 215)			
Child Assessments	98	95	90
Teacher Report	97	95	82
Parents Interview	91	94	86
University of Texas-Houston (n = 297)			
Child Assessments	99	94	79
Teacher Report	97	86	57
Parents Interview	80	74	68
University of North Florida (n = 244)			
Child Assessments	100	92	89
Teacher Report	96	89	64
Parents Interview	84	81	73
University of Virginia (n = 195)			
Child Assessments	85	96	97
Teacher Report	87	93	81
Parents Interview	93	87	89
Florida State University (n = 297)			
Child Assessments	95	96	80
Teacher Report	96	93	80
Parents Interview	91	84	75
UC-Berkeley and University at Buffalo, SUNY (n = 316)			
Child Assessments	99	94	90
Teacher Report	99	94	74
Parents Interview	83	90	78
Purdue and University of WI-Milwaukee (n = 204)			
Child Assessments	100	94	85
Teacher Report	100	90	66
Parents Interview	86	76	70

See notes at end of table.

Table B-3. Response rates and attrition—Continued

Research team	Response rate Fall 2003	Percent of sample with data Spring 2004	Percent of sample with data Spring 2005
University of Missouri-Columbia (n = 231)			
Child Assessments	99	90	81
Teacher Report	98	81	68
Parents Interview	92	84	84
UC-Berkeley (n = 286)			
Child Assessments	96	92	87
Teacher Report	96	95	84
Parents Interview	91	82	76
All teams (n = 2,911)			
Child Assessments	98	93	85
Teacher Report	97	90	72
Parents Interview	84	79	75

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

Missingness—Item Non-response

To address item non-response in the parental interviews, we created missing status indicators for each of the parent interview categorical variables, to keep these cases in the analyses. This approach allowed us to fit a separate categorical effect for non-responders. The parameter estimates for the predictor variable associated with each missing data dummy variable is estimated based on available cases. This approach allowed for the inclusion of the maximum number of participants in all analyses and estimated associations between the predictors and outcomes based on all available cases.

Overall Modeling Framework

The general framework in which all of the analyses were conducted was the linear mixed effects model. This model generalizes the more familiar general linear model for regression by allowing for both random and fixed regression coefficients. The data analyses were conducted at the research team level. This was an appropriate approach because the selection and randomization of classrooms to treatment versus control condition were completed at the grantee site level for each intervention curriculum. Children were nested within classrooms in both the intervention and control groups and were repeatedly assessed with a battery of measures. These various sources of nesting (time within child and children within classrooms) were accounted for in the analyses. A hierarchical linear model (HLM) was therefore used for all study analyses to deal with the correlated (i.e., nested and repeated measures) data as well as the mixture of random and fixed effects in each model. The HLM model more accurately accounts for the variance in the data. This is important in order to obtain valid variance estimates and significance tests.

For child-outcome models, classroom was included as a random factor to account for the design effects associated with nesting of children within classrooms. We elected to test all treatment versus control inferences at a level generalizable to the entire population of similar students and classrooms/programs, which has the effect of increasing standard errors over inferences made only to the specific instances included in the study. Models that explicitly modeled change over time (i.e., repeated measures and repeated measures spline models) included random effects for "time."

The underlying assumptions for this model have been met, within reasonable limits. The main assumption is that each measure modeled is distributed according to a multivariate normal distribution. The linear mixed-effects model is relatively robust to some non-normality, such that extreme multivariate skewness or kurtosis would be needed to invalidate the model. It also assumes a linear relationship between the predictors and the outcomes. The model deals with the lack of independence among children in the same classroom by the inclusion of the random classroom intercept. The lack of independence due to repeated measures from the same children is addressed in the model by the inclusion of a covariance matrix for the errors that model this dependence among repeated measures.

Overview of Quantitative Modeling Utilized in the Study's Primary and Secondary Analysis

As explained in the main report,[2] analysis on nearly all outcome measures was conducted in two different ways, one of which was considered primary and reported on in the main report, the other was considered secondary and was reported on only in appendix A. Within these two analysis approaches, models varied depending on the data structure of the longitudinal data (i.e., how many and at which time points data for a given outcome were collected). Tables B-4 and B-5 provide a summary overview of all the primary and secondary analyses presented in the main report and in appendix A.

Main Analysis—Reported in the Main Body of the Report

In the main report, we present repeated measures spline model analyses for data collected at three time points, simple repeated measures analyses for data collected at two time points, and ANCOVA analyses for data collected at one time point. We discuss these analyses below beginning with the repeated measures spline model, followed by a description of the simple repeated measures and ANCOVA analyses.

Secondary Analysis—Reported in Appendix A

In appendix A, we present additional analysis results from the repeated measures spline model analyses of data from three time points, analysis results from the simple repeated measures analyses of data from two time points, and ANCOVA analysis results of data from two or three time points. These secondary analyses were conducted to address questions related to nonequivalence at baseline and the possibility of early treatment effects.

[2] The "main report" refers to chapters 1-13.

Table B-4. Main analysis: Model used with each outcome measure

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

[1] Pre-kindergarten and kindergarten measures not on the same scale.
[2] Building Blocks, Shape Composition task
NOTE: ANCOVA: Analysis of covariance. The repeated measures spline model was used to analyze data collected at three time points (fall and spring of pre-kindergarten and spring of kindergarten). The simple repeated measures model was used to analyze data collected at two time points (fall and spring of pre-kindergarten). Refer to the glossary for abbreviations of the measures.
SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Table B-5. Secondary analysis: Models and grades for each outcome measure

Outcome	Measure	Repeated Measures model	ANCOVA model with Pre-K baseline covariate
Reading	TERA	Spline: Pre-K and K	Pre-K and K
	WJ Letter Word Identification	Spline: Pre-K and K	Pre-K and K
	WJ Spelling	Spline: Pre-K and K	Pre-K and K
Phonological awareness[1]	Pre-CTOPPP	Simple: Pre-K	Pre-K
	CTOPP		K
Language	PPVT	Spline: Pre-K and K	Pre-K and K
	TOLD	Spline: Pre-K and K	Pre-K and K
Mathematics	WJ Applied Problems	Spline: Pre-K and K	Pre-K and K
	CMA-A Mathematics Composite	Spline: Pre-K and K	Pre-K and K
	Shape Composition[2]	Spline: Pre-K and K	Pre-K and K
Pre-kindergarten behavior[1]	SSRS Social Skills	Simple: Pre-K	Pre-K
	SSRS Problem Behaviors	Simple: Pre-K	Pre-K
	PLBS	Simple: Pre-K	Pre-K
Kindergarten behavior[1]	SSRS Social Skills		K
	SSRS Problem Behaviors		K
	LBS		K
Classroom quality	ECERS-R	Simple: Pre-K	Pre-K
Teacher-child interaction	Arnett Detachment	Simple: Pre-K	Pre-K
	Arnett Harshness	Simple: Pre-K	Pre-K
	Arnett Permissiveness	Simple: Pre-K	Pre-K
	Arnett Positive Interaction	Simple: Pre-K	Pre-K

[1] Pre-kindergarten and kindergarten measures not on the same scale.
[2] Building Blocks, Shape Composition task
NOTE: ANCOVA: Analysis of covariance. The repeated measures spline model was used to analyze data collected at three time points (fall and spring of pre-kindergarten and spring of kindergarten). The simple repeated measures model was used to analyze data collected at two time points (fall and spring of pre-kindergarten). Refer to the glossary for abbreviations of the measures.
SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Data Analysis Models

Repeated Measures

The repeated measures model provides a measure of intervention effects that is less influenced by differences in the variables under investigation that might exist between the two groups at baseline. Also, by including Time, the length of exposure to intervention, as a covariate, the model results (e.g., estimates of intervention group means or group differences at, for example, the pre-kindergarten spring time point) are less influenced by an early treatment effect that might exist at some sites. The inclusion of time also provides estimates of treatment and control group rates of change, allowing the differences in their rates of change to be tested.

Repeated Measures Linear Spline Model

For data collected in the fall and spring of pre-kindergarten and the spring of kindergarten, a repeated measures linear spline model was applied that modeled the three scores, using Time variables and a set of covariates. A simple graphical display of this model, also known as a "piece-wise linear" or "broken stick" model, is found in figure B-1. As the figure indicates, this model allows projection of the intervention and control group trajectories back to the start of the school year when curriculum implementation began. Group differences at the start of treatment can then be tested, assuming straight-line growth throughout the pre-kindergarten year.

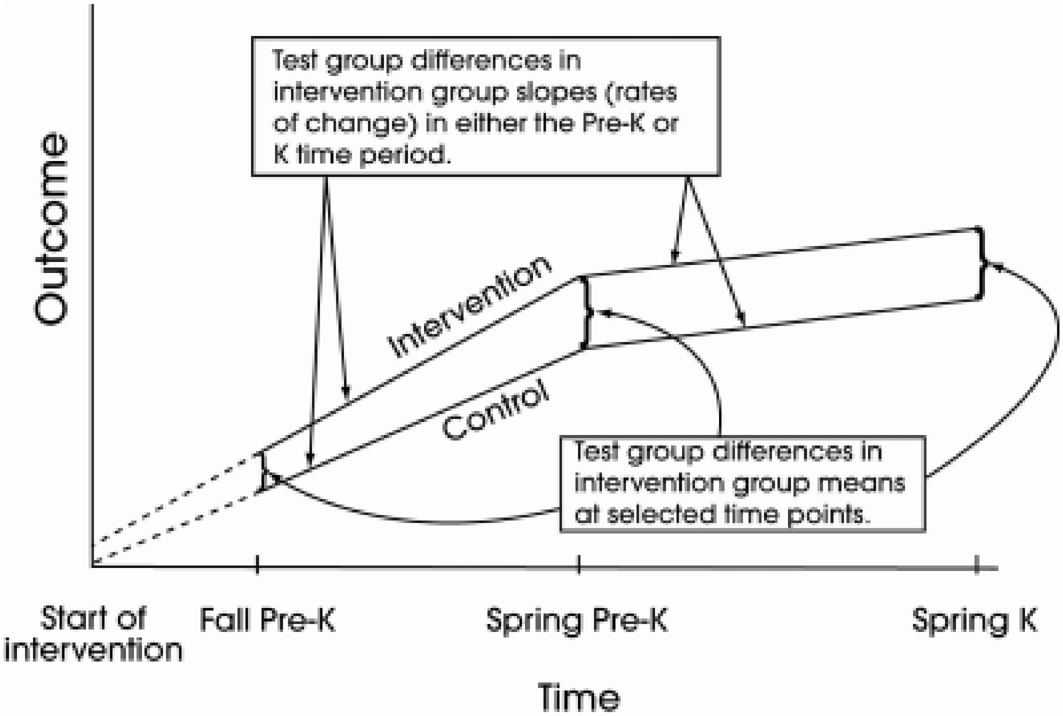
This model allows for the most extensive testing of the first two research questions: Question 1, regarding the child impacts at the end of preschool, and Question 2, regarding the child impacts at the end of kindergarten (some of which were sustained from preschool and others evident only at the end of kindergarten), and provides the most complete account of the data. Intervention-control group differences can be obtained at many points of interest: the spring pre-kindergarten and spring kindergarten time points, as well as baseline and start-of treatment time points. Group differences in the rate of change between intervention and control groups can be tested during the pre-kindergarten and kindergarten year (testing for "sleeper" and maintenance of intervention effects), as can within-group comparisons of the rates of change between the 2 years.

Simple Repeated Measures

For the four child outcomes and two classroom outcomes with comparable data from two time points, simple repeated measures analyses were conducted. This model accounts for repeated measures data by fitting a linear growth trajectory using the fall and spring time point means on the outcome of interest, adjusted for the covariates included in the model. Figure B-2 graphically displays this simple repeated measures model. As illustrated, the pre-kindergarten growth trajectories estimated by this model can be extended back to the start of treatment, allowing for a test of group differences at the start of treatment.

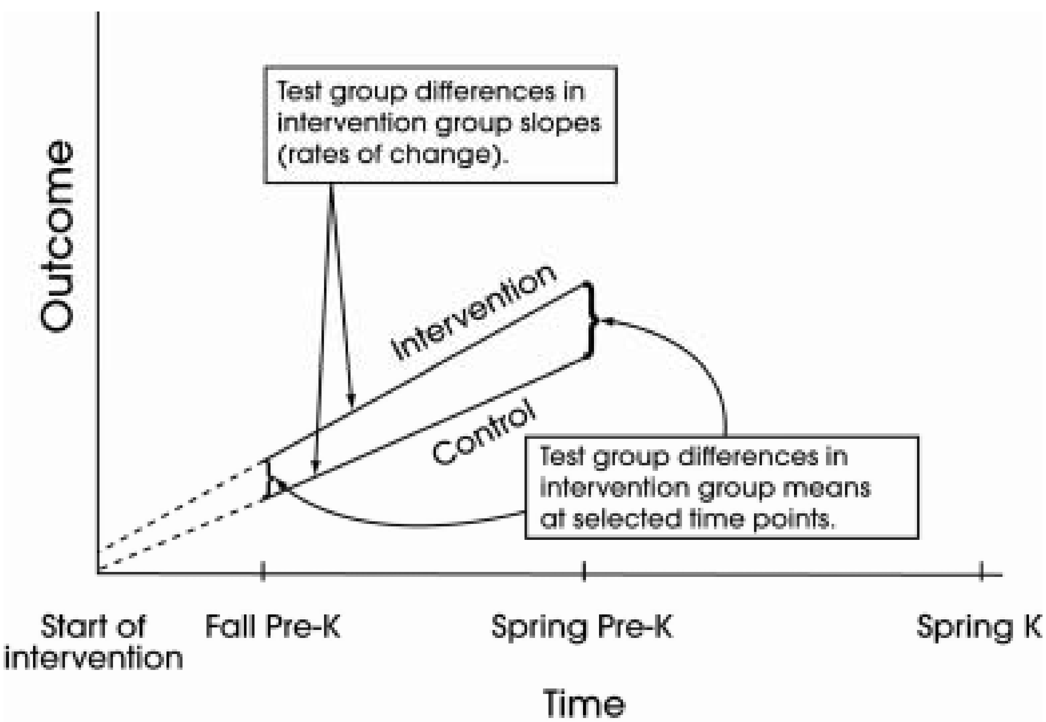
This model allows testing for intervention differences in spring scores adjusted for the covariates *and* for the amount of exposure to the intervention, as reflected in the Time variable. The model also affords parameter estimates to test the primary child research Question 1 (child outcomes at the end of pre-kindergarten) and research Question 3 (classroom outcomes at the end of pre-kindergarten). This model is similar to the repeated measures spline model described above, except that with two time points, it only models a single segment of the outcome trajectories covering the pre-kindergarten year.

Figure B-1. Repeated measures spline model



NOTE: Pre-K (pre-kindergarten); K (kindergarten)
SOURCE: The Preschool Curriculum Evaluation Research (PCER) study.

Figure B-2. Simple repeated measures model



NOTE: Pre-K (pre-kindergarten); K (kindergarten)
SOURCE: The Preschool Curriculum Evaluation Research (PCER) study.

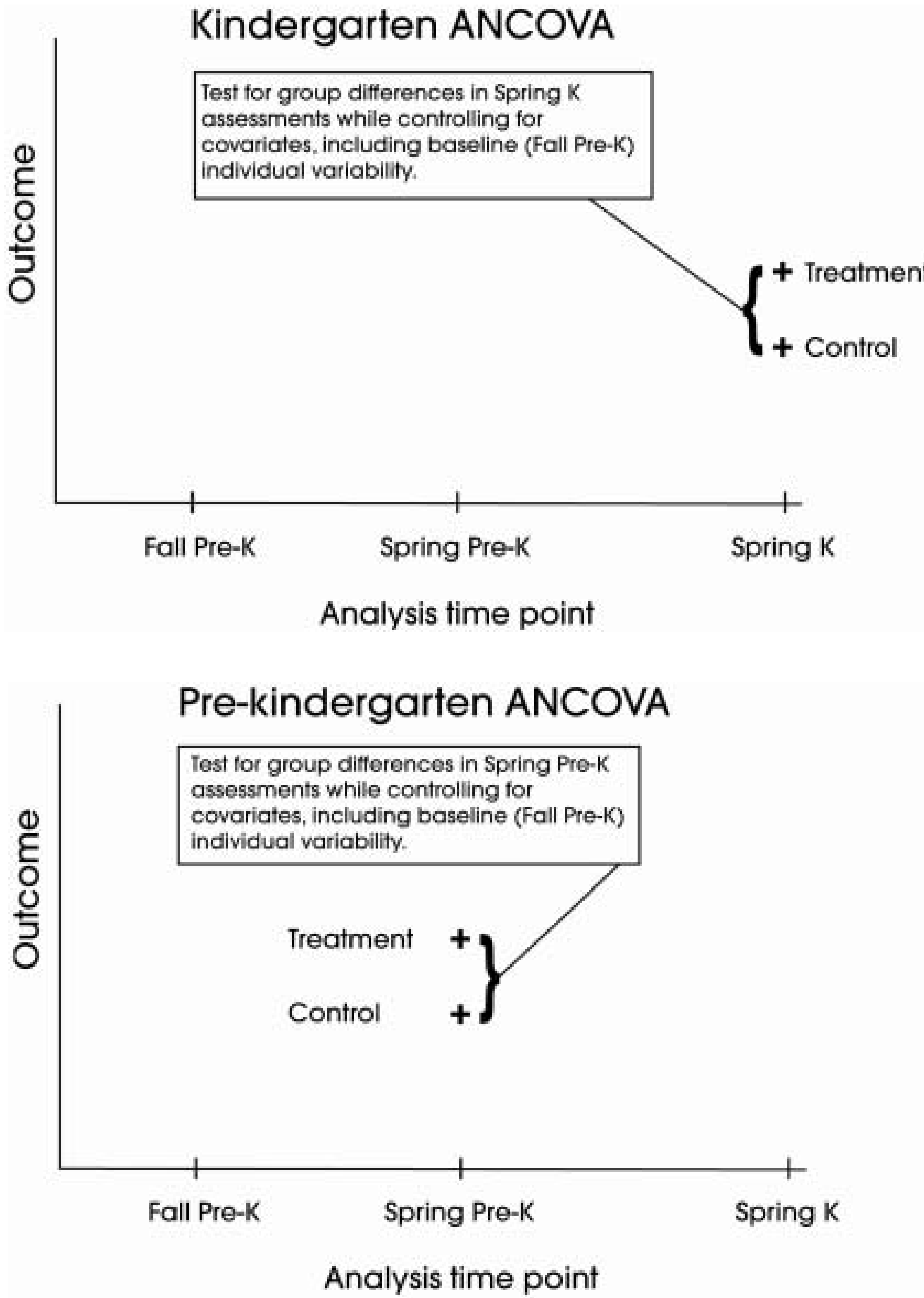
ANCOVA Model

For findings presented in the main report, an ANCOVA model was used to analyze comparable child or classroom data from one time point (see table B-4). In the ANCOVA analyses of child outcomes, the pre-kindergarten baseline score and the child and family covariates were included in the analysis model. For the ANCOVA analysis of classroom outcomes (TBRS classroom observation measure), the teacher, preschool, and community covariates were included in the model.

In appendix A, we present ANCOVA analyses for data from all time points. Two ANCOVA models, each modeling one or two spring outcomes using the child's fall pre-kindergarten score as a covariate, along with the child and family demographic covariates, were used to analyze child outcomes for which we have data from one, two, or three time points. We also present ANCOVA analyses for classroom outcomes with comparable data from one or two time points in appendix A.

Figure B-3 provides a graphical presentation of the ANCOVA model used with individual data for the child outcomes, and classroom data for the classroom outcomes. The ANCOVA model allows modeling of spring outcomes for both years, in spite of possible changes in measurement between either spring outcome and the baseline covariate. The ANCOVA model has a couple of disadvantages, namely, that (a) no rates of change can be estimated because the baseline is not modeled, and (b) this model may be biased by the possibility of an early treatment effect due to late baseline data collection, because the baseline assessment score is included in the model as a covariate. This model only provides estimates and tests of significance for group differences at the spring time point (end of pre-kindergarten or end of kindergarten) in each model. However, this is sufficient to provide answers to research Questions 1 and 2 for the child outcomes, and Question 3 for the classroom outcomes examined with this model.

Figure B-3. Pre-kindergarten (Pre-K) and kindergarten (K) analysis of covariance (ANCOVA) models



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

Statistical Models

The analyses were conducted using the general mixed model framework. This model generalizes the general linear model (GLM) to include random as well as fixed effects, and allows for a more general variance-covariance structure for the covariates as well as error structure. Repeated measures and ANCOVA forms of the general mixed model were used to analyze the pre-kindergarten and kindergarten data.

General Multi-Level Model

A multi-level model (mixed-effect model) can be written in a matrix format as

$$\underset{\sim}{Y} = X\underset{\sim}{\beta} + Z\underset{\sim}{u} + \underset{\sim}{\varepsilon} \qquad (1)$$

where

$\underset{\sim}{Y}$ is the n-dimension vector of observations, with n being the number of subjects in the study

$X_{n \times p}$ is the $n \times p$ design matrix for fixed effects, with *p* being the number of fixed effects

$\underset{\sim}{\beta}_p$ is the p-dimension vector of fixed effect coefficients

$Z_{n \times r}$ is the $n \times r$ design matrix for the random effects, with r being the number of random effect parameters

$\underset{\sim}{u}_r$ is one vector of random effect parameters

$\underset{\sim}{\varepsilon}_n$ is the n-dimension residual random error.

In this model, everything is the same as in the general linear model except for the addition of the known design matrix, Z , and the vector of unknown *random-effects parameters,* y . The matrix Z can contain either continuous or dummy variables, just like X . The name *mixed model* comes from the fact that the model contains fixed effects parameters, β, and random-effects parameters, $\underset{\sim}{u}$. Henderson (1990) and Searle, Casella, and McCulloch (1992) provide a discussion of the historical developments of the mixed model.

If the covariance of the vector of random effects and the vector of error terms are given by

$$V(\underset{\sim}{u}) = G \text{ and } \mathrm{V}(\underset{\sim}{\varepsilon}) = R \qquad (2)$$

respectively, then the overall covariance structure of the observations is given as

$$V(\underset{\sim}{Y}) = ZGZ' + R \qquad (3)$$

A key assumption in the foregoing analysis is that $\underset{\sim}{u}$ and ε are normally distributed with

$$E\begin{bmatrix} \underset{\sim}{u} \\ \underset{\sim}{\varepsilon} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
$$Var\begin{bmatrix} \underset{\sim}{u} \\ \underset{\sim}{\varepsilon} \end{bmatrix} = \begin{bmatrix} G & \Theta \\ \Theta & R \end{bmatrix} \qquad (4)$$

One can obtain $V(\underset{\sim}{Y})$ once both the random-effects design matrix Z and estimates from specified covariance structures for G and R are obtained. The covariance structures that can be used include variance components, unstructured, compound symmetry, and various time-dependent structures among others.

Repeated measures models

A repeated measures model, which is just a special case of the general model (1), was used to model assessments from the pre-kindergarten and kindergarten years simultaneously.

As an example, a model with only a random intercept term, if written in scalar format, is specified as

$$y_{ijk} = X_{ijk} \underset{\sim}{\beta} + u_i + \varepsilon_{ijk}, \; i = 1, 2, \cdots, r, \quad j = 1, 2, \cdots, n_i, \quad k = 1, 2, 3$$

where r is the number of classrooms, n_i is the size of classroom i, (ij) indicates the j^{th} individual in the i^{th} classroom, u_i is the random intercept term, and k indicates the time point of the assessment (fall pre-kindergarten, spring pre-kindergarten, or spring kindergarten). X_{ijk} may include a time effect, an intervention grouping variable, an interaction between intervention group and Time, along with other covariates. The Time variable was defined as the time since start of treatment for any given piece of assessment data. For the repeated measures linear spline model, X_{ijk} included two Time variables ($Time_1$ and $Time_2$). The first Time variable, $Time_1$, is the time since start of treatment. The second Time variable, $Time_2$, is the time since the pre-kindergarten spring assessment to the kindergarten assessment, and was defined as 0 (zero) for the pre-kindergarten assessments. The $Time_2$ variable provides a means of estimating any kindergarten-year increase or reduction in slope over and against the earlier pre-kindergarten-year slope that is represented by the coefficient for $Time_1$. The spline model included these two Time variables along with group interactions with each of the two Time variables: Intervention Group x $Time_1$ (for the pre-kindergarten period) and Intervention Group x $Time_2$ (for the kindergarten period). These interaction terms allowed us to include group-specific slopes for both time periods covered by the model.

When there was more than one random effect in the model (i.e., random intercepts and random slopes), a covariance structure among the random effects needed to be specified. Models treating $Time_1$ and/or $Time_2$ as random, in addition to the classroom intercept, were tested. In only one case was a random Time variable kept in the model (see table B-14, column regarding random variables). All grouping variables, covariates, and their interactions that were not treated as random were treated as fixed effects in the model.

Each child contributed three observations to the dataset used for model fitting. We imposed a correlation structure on observations from the same subject. Several different structures may be imposed, and we used a variety of structures to model this as appropriate (see Covariance Structure and Modeling Steps section for a description of the specific structures used in our analysis).

ANCOVA model

In the ANCOVA model for child outcomes, the fall pre-kindergarten (or baseline) measure was included in the model as a covariate, to increase precision by accounting for the variance that can be attributed to possible differences in children's scores on the baseline measures even with random assignment of groups. The child and family demographic covariates (i.e., child's age, race/ethnicity, gender, disability status, and maternal education) were also included in the model to increase precision of the estimates by accounting for any possible baseline differences on these characteristics. The model can be written in the form of equation (1), with one of the columns of X being the baseline score. Another column in X would be the site-specific or curricula-within-grantee-specific grouping variable, which captures the intervention impact. As an example, the following is a model used with a single random intercept and a correlated error term:

$$Y_{ij,t} = \beta_0 + u_i + \beta_1 Y_{ij,\text{baseline}} + X\beta + \varepsilon_{ij} \qquad (5)$$

for the full model with covariates, where $Y_{ij,t}$ is the score at time t (say, spring kindergarten), $Y_{ij,baseline}$ is the fall pre-kindergarten score, $ui \sim N(0, \sigma_{u^2})$, and $\varepsilon \sim N(0, R)$, for $j = 1$ to n_i individuals, and $i = 1$ to r classrooms.

The effects of interest in this analysis are the differences between the intervention and control group means. We used ESTIMATE statements in SAS PROC MIXED statistical software to provide adjusted means and two-tailed t-tests of intervention and control group mean differences at the start of intervention, fall, and spring assessments.

The confound between site and curriculum was handled by effect coding "Curriculum-in-Grantee" or "Site-in-Grantee." Thus, we created separate variables for each curriculum at each site (e.g., *Pre-K Mathematics with DLM Early Childhood Express Math software package* in California, *Pre-K Mathematics with DLM Early Childhood Express Math software package* in New York, *Creative Curriculum* in North Carolina and Georgia) and coded control classrooms as -.5 and intervention classrooms as .5. Florida-FSU and the University of Texas-Houston research teams each had two intervention conditions that shared a control group; their intervention and control conditions were coded 1 and -1, respectively. Simple linear estimates and contrasts were used to specify model-predicted means at the fall and spring data collection points as well as model-predicted gain scores. This approach provided a parsimonious way of representing the design of the site-specific projects (e.g., a site with one intervention curriculum and one control group or a site with two intervention curricula and one control group).

Structures of Residual Covariance Matrix R

The covariance matrix of the residual errors for the repeated measures models are defined at the individual level, and takes the form of a block diagonal matrix. If we use R_{ij} to denote the covariance among observations from the j^{th} child in the i^{th} classroom, the $(i,j)^{th}$ diagonal block of R is defined for each model differently depending on the number of observations and the covariance structure we chose for the model. The notation (i,j) is not the typical $(i,j)^{th}$ element of a matrix, in this instance, all R_{ij} 's are diagonal blocks of R. The residual covariance matrix is represented in the following form:

$$R = \begin{pmatrix} R_{n_1} & \cdots & \Theta \\ \vdots & \ddots & \vdots \\ \Theta & \cdots & R_{n_r} \end{pmatrix} \quad \text{and} \quad R_{n_1} = \begin{pmatrix} R_{i,1} & \cdots & \Theta \\ \vdots & \ddots & \vdots \\ \Theta & \cdots & R_{i,n_i} \end{pmatrix}$$

For the linear spline models with three time points, a heterogeneous compound symmetry (HCS) structure that assumes the covariance among residual errors of the same subject was used:

$$Cov\left(\underset{\sim}{\varepsilon}_{ij}\right) = R_{i,j} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 & \sigma_2\sigma_3\rho \\ \sigma_1\sigma_3\rho & \sigma_2\sigma_3\rho & \sigma_3^2 \end{pmatrix} \qquad \text{for all } i, j$$

For the repeated measures models for data from two time points, an unstructured covariance structure was used and defined as

$$Cov\left(\underset{\sim}{\varepsilon}_{ij}\right) = R_{i,j} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \qquad \text{for all } i, j$$

For repeated measures models of the classroom outcomes (for data with two time points), a variance component covariance structure was used and defined as

$$Cov\left(\underset{\sim}{\varepsilon}_{ij}\right) = R_{i,j} = \begin{pmatrix} \sigma_0^2 & 0 \\ 0 & \sigma_0^2 \end{pmatrix} \qquad \text{for all } i, j$$

Structures of Random Effect Covariance Matrix G

For the linear spline models with three repeated measures, after experimenting with different structures for the random effects (including both the random intercept and slopes), a model with only a random intercept term was settled on:

$$G = \sigma_u^2 I \text{, and } Z = \begin{bmatrix} \mathbf{1}_{n_1} & & & \\ & \mathbf{1}_{n_2} & & \\ & & \ddots & \\ & & & \mathbf{1}_{n_r} \end{bmatrix}$$

where $\mathbf{1}_{n_i}$ is a vector of 1's of dimension n_i.

When modeling child outcomes with comparable data from two observations (i.e., behavioral outcome models, see Covariance Structure and Modeling Steps for more details), a random slope seemed justified. Therefore, those models include a random slope term, as well a random intercept term. The structure of G and Z are more complicated and defined as

$$G = \begin{pmatrix} \Sigma & \cdots & \Theta \\ \vdots & \ddots & \vdots \\ \Theta & \cdots & \Sigma \end{pmatrix} \text{ and } Z = \begin{bmatrix} Z_{n_1} & & & \\ & Z_{n_2} & & \\ & & \ddots & \\ & & & Z_{n_r} \end{bmatrix}$$

where G and Z are a block diagonal matrices of the same dimensions. The diagonal blocks of G are identical and are given as

$$\Sigma = \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_s^2 \end{pmatrix}$$

with σ_u^2, and σ_s^2 being the variance components for the random intercept and random slope, respectively, and Zn_i being matrices of size $2 \times n_i$ with the first column containing all 1's and second column containing the observed values of Time (time interval between the start of intervention and the particular follow-up assessment) for each subject in classroom i:

$$\begin{bmatrix} 1 & \mathrm{T}_{i,1} \\ \vdots & \vdots \\ 1 & \mathrm{T}_{i,n_i} \end{bmatrix}$$

Assumed Variance Structure of Outcomes

After making assumptions regarding covariance matrices G and R, the assumed overall covariance structure of the outcome under modeling can be derived from the following relationship:

$$V\left(\underset{\sim}{Y}\right) = ZGZ' + R$$

As an example, the assumed covariance structure for the repeated measures models with both a random intercept, a random slope, and observations from two time points is given as

$$Cov\left(y_{ijk}, y_{i'j'k'}\right) = \begin{cases} \sigma_u^2 + T_{i,j}^2\sigma_s^2 + \sigma_1^2 & \text{if } i = i', j = j', k = k' = 1 \\ \sigma_u^2 + T_{i,j}^2\sigma_s^2 + \sigma_2^2 & \text{if } i = i', j = j', k = k' = 2 \\ \sigma_u^2 + T_{i,j}^2\sigma_s^2 + \sigma_{12} & \text{if } i = i', j = j', k \neq k' \\ \sigma_u^2 + T_{i,j}T_{i,j'}\sigma_s^2 & \text{if } i = i', j \neq j' \\ 0 & \text{if } i \neq i' \end{cases}$$

where i is the index for classroom, j for individual child, and k for time point. The structure implies that observations from children of different classrooms are independent. Observations from children of the same classroom are correlated and the covariance is accounted for by the intra-classroom covariance σ_u^2 and the covariance of the random classroom slope σ_s^2. Observations from the same child are correlated, and the covariance depends on both the intra-classroom correlation and the correlation among observations of the same child (as specified in the residual covariance matrix R).

Mean Model and Testing of Fixed Effects

To illustrate this mixed model approach we use a slightly simplified linear spline model (excluding covariates) with two groups (intervention and control). Two fully developed examples from our analyses are found in the section Group Comparisons Testing for Intervention Impact. The model with the Time_1, Time_2, Group, and their attendant interaction terms is as follows:

$$E\left(Y_{ij}\right) = \beta_1 + \beta_2\mathrm{Time}_1 + \beta_3\mathrm{Time}_2 + \beta_4\mathrm{Group}_i + \beta_5\left(\mathrm{Time}_1 \times \mathrm{Group}_i\right) + \beta_6\left(\mathrm{Time}_2 \times \mathrm{Group}_i\right)$$

For this model, when the control group is coded as 0 (zero), its expectation is

$$E\left(Y_{ij}\right) = \beta_1 + \beta_2\mathrm{Time}_1 + \beta_3\mathrm{Time}_2$$

whereas the intervention group which is coded as one, has as its expectation:

$$E\left(Y_{ij}\right) = \left(\beta_1 + \beta_4\right) + \left(\beta_2 + \beta_5\right)\mathrm{Time}_1 + \left(\beta_3 + \beta_6\right)\mathrm{Time}_2$$

Hypotheses tested included:

$H_0 : \beta_4 = 0$ (No group difference at start of intervention),

$H_0 : \beta_5 = 0$ (No group difference in pre-kindergarten growth rate),

$H_0 : \beta_5 + \beta_6 = 0$ (No group difference in kindergarten growth rate),

$H_0 : \beta_3 = \beta_6 = 0$ (No change in the slope in the kindergarten year).

In the analyses, effect coding is used rather than dummy coding; therefore, the betas have a somewhat different meaning than that illustrated above. With effect coding, the control group is coded -1 and the treatment group is coded 1 (or some variant of this coding) to account for the various grantee design configurations. This coding scheme results in the regression parameters being equal to treatment effects obtained in traditional analysis of variance models. For example, with effect coding, β_1 represents the overall mean, and β_4 represents the effect of the treatment relative to the control group (Kirk 1995). As with the example provided above, the contrasts to test for intervention/control group differences in mean levels and slopes were all estimable and conducted with the effect coding used in the model specification.

Sample Weights and Missing Data

Sample weighting to account for variability across sites in sample sizes and intervention-control balance is unnecessary in the mixed model. The mixed model approach used for the ANCOVA or repeated measures spline model accounts for the unbalanced design when analyses are conducted using SAS PROC MIXED. The PROC MIXED analysis is valid with unbalanced data and missing data, as long as they are missing randomly (Littell et al. 1996). No contrary evidence was found to this assumption.

Estimation Methods

As indicated above, the mixed model that was used must estimate parameters in four matrices: (1) the variance-covariance matrix for the random effects, G; (2) the error structure for the residual error, R; (3) the fixed effects parameters, β; and (4) the random-effects parameters, y. The elements of G and R are estimated using restricted, or residual, maximum likelihood (REML). Fixed parameters are estimated using the Maximum Likelihood (ML) method. SAS PROC MIXED was used to carry out the estimation. PROC MIXED allows for three types of estimation: ML, REML, or MIVQUE0 (Minimum Variance Quadratic Unbiased Estimation). Swallow and Monahan (1984) showed that REML and ML variance component estimates are superior to MIVQUE0. As reported in Swallow and Monahan (1984), under some circumstances, MIVQUE0 is a poor estimator of unbalanced data. They also note that the ML estimates have the smallest MSE (means square error); however, this is due to the downward bias in these estimates. The REML-estimated MSEs are larger precisely because the estimates are unbiased. REML estimates are unbiased and considered superior to ML estimates because they account for the degrees of freedom lost in the estimation of the fixed effects. We further discuss the relative merits of REML and ML estimation in the next section. The Proc Mixed estimation method was implemented using a sweep-based Newton-Raphson algorithm (Wolfinger et al. 1994).

For the fixed model estimates, you have to obtain estimates of β and y the standard method is to solve the mixed model equations (Henderson 1984):

$$\begin{bmatrix} X'\hat{R}^{-1}X & X'\hat{R}^{-1}Z \\ Z'\hat{R}^{-1}X & Z'\hat{R}^{-1}Z + \hat{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} X'\hat{R}^{-1}y \\ Z'\hat{R}^{-1}y \end{bmatrix} \tag{6}$$

The solutions can also be written as

$$\hat{\beta} = \left(X'\hat{V}^{-1}X \right) - X'\hat{V}^{-1}y$$
$$\hat{\gamma} = \hat{G}Z'\hat{V}^{-1}\left(y - X\hat{\beta} \right) \tag{7}$$

and have connections with empirical Bayes estimators (Laird and Ware 1982). These estimates are obtained in SAS PROC MIXED using maximum likelihood (ML) estimation.

When estimates for G and R are used to estimate β and y, the resulting estimates have favorable statistical properties. The estimate of β is considered to be the empirical best linear unbiased estimator (EBLUE) of β, and the estimate of y is considered an empirical best linear unbiased predictor (EBLUP) of y.

REML Versus ML Estimation Method

SAS PROC MIXED offers two methods of maximum likelihood estimation of the variance components. One is the full maximum likelihood (FML or ML), and the other is the restricted (or residual) maximum likelihood (REML or RML). Both use the likelihood principle but differ with regard to how the likelihood function is constructed.

The full ML estimates of the variance components contain ML estimates of the fixed effects when estimating variance components. This method treats the estimated fixed effects as known. By ignoring the uncertainty due to the fact that the fixed effects are sample estimates, ML estimates of the variance components overstate the degrees of freedom available for estimation and therefore underestimate the variance of the fixed effect parameters. Such concerns led to the development of REML (Patterson and Thompson 1971; Harville 1974; Dempster, Laird, and Rubin 1977). REML maximizes a residual version of the likelihood function, after the estimation of the fixed effects. As a result, the variance component estimates obtained from the two approaches can be numerically different. The point can be illustrated using an example of an ordinary regression model (with no random effects), assuming we have a total of N independent observations and if we fit an ordinary regression model with p parameters (fixed effects). If we further denote the true variance of the observations as σ^2, then the ML and the REML versions of the estimate of variance have the following relationship with the true variance:

$$E(\hat{\sigma}^2_{ML}) = \frac{(N-p)}{N}\sigma^2, \quad \text{and} \quad E(\hat{\sigma}^2_{REML}) = \sigma^2 \tag{8}$$

Thus, the REML estimate of the variance is unbiased and adjusts the ML estimate of variance by the following multiplier:

$$\frac{N}{N-p} = \frac{\text{Number of observations}}{\text{Number of observations} - \text{Number of fixed effect parameters}} \geq 1 \tag{9}$$

For random effect models, both ML and REML estimates are calculated through iterative procedures and an equivalent adjustment is made. The REML likelihood function is obtained by multiplying the usual ML likelihood function by a factor that is the square root of the generalized variance of the fixed effect parameters.

To understand the impact of this adjustment, we again use the above example of ordinary regression model. The multiplier defined in (9) is close to 1 when the number of fixed parameters is much smaller than the number of observations. However, when the difference of the two is small, it could be quite different from one. The number of fixed effect parameters used in our models differs by model; however, they are the same for a given model across all sites and outcomes. For example, the linear spline model described in the section on the repeated measures models has 21 fixed effect parameters and the ANCOVA model described in the section ANCOVA model has 16 fixed effect parameters. The number of observations used in fitting those models to child outcomes ranges between 300 and 100, with the majority in the lower portion of the range nearer 100. The smaller sites, however, had sample sizes as low as 50 (NH ANCOVA model). For the classroom outcomes models, the number of observations included in the models is even smaller. As a result, the adjustment factor (9) can sometimes be substantially larger than 1. That is, ML estimates can underestimate the true value by a substantial amount.

Although REML is supposed to be less biased than ML when sample sizes are small, it is not guaranteed to produce numerically superior estimates in all circumstances (Kreft and deLeeuw 1998). One reason for this is that REML estimates may become more variable than ML estimates as the variance of the fixed effect parameters becomes larger (or the sample size becomes smaller), thus increasing the possibility of REML behaving worse than ML. Again, using the ordinary regression model in (8) as an example, if $N = 30$ and $p = 20$, then REML estimates would have degrees of freedom of 10 versus 30 for the ML estimates. As a result, although the expected REML estimates are still unbiased, the probability of the REML estimate of a particular sample being further away from the true value of the parameter is much higher than the corresponding ML estimate.

As one cannot know for sure whether this less desirable property of REML is being manifested in any particular situation, the odds of that occurring must be balanced against the almost certain bias that occurs with ML estimates. The consensus is that REML is theoretically the approach of choice. Dempster, Laird, and Rubin (1977) declared REML to be "intuitively more correct." In addition, REML solutions have the desired property of being equivalent to the ANOVA estimators for balanced models. Most standard texts on mixed models (Verbeke and Molenberghs 2000; McCulloch and Searle 2001; Fitzmaurice, Laird, and Ware 2004) advise using REML, and consider it the preferred approach. As a result, we used REML estimates for all reported significance levels.

Early Intervention Effects and Estimated True Baseline

The start of baseline data collection began 2 weeks or more after the start of the school year (start of intervention curriculum implementation) at several research sites. This data collection timeline raised concerns among members of the PCER Consortium that early effects of the experimental curriculum may have already occurred by the time the baseline assessments were collected. Failure to obtain true baselines (by which we mean measures of child and classroom data uncontaminated by the start of intervention) may result in biased estimated growth rates and growth rate impacts during the pre-kindergarten year, with the direction of that bias undetermined. It is difficult to ascertain whether this situation occurred because we did not collect data from children prior to the start of the school year.

To address concerns related to early treatment effects due to the absence of a "true" baseline, we included a time variable that reflects the amount of exposure to the intervention in the repeated measures models (see *section on* Repeated Measures Models). This time variable takes a value of 0 (zero) at the initiation of the intervention and measures all other assessment time points from this origin. Although the estimated baseline group means may still have been affected by early intervention effects, with this covariate in the model, the estimated rate of change from repeated measure models is much less influenced by potential early treatment effect due to delays in the baseline assessment.

Using this time variable, group means were estimated at time 0 (zero) (i.e., the start of the intervention) for both the intervention and control groups even when the baseline assessment was conducted weeks later. These means were estimated the same way the group means were estimated for other time points in the model. The means were estimated by just setting $Time_1$ and $Time_2$ equal to 0 (zero) in the spline model or just Time = 0 in the simple repeated measures model. Significance testing of the extrapolated group differences at time 0 (zero) allows us to conclude, under the model assumptions, whether the groups differed at the start of intervention (beginning of the school), not just at baseline (time of the fall pre-kindergarten assessment). This set of estimated means is not problem-free, as it relies upon an assumption that the growth process is relatively linear from start of intervention through the spring assessment. To the extent this assumption does not hold, particularly in the period of time before the baseline assessment, the start-of-intervention estimates should be interpreted with caution. Despite the possible limitations, extrapolating back to the start of the intervention and estimating treatment and control groups mean differences for the start of treatment and fall baseline assessment was our attempt to address issues related to the possibility of early treatment effects.

The results across all 14 curricula indicated that there were very few instances in which there were statistically significant differences between the intervention and control groups at start of intervention. This suggests, if the linear growth assumption holds, that there is very little evidence of imbalance between the intervention and control groups in terms of the child outcomes at the start of the study, after adjusting for the standard set of covariates. However, in some cases (e.g., classroom-level models for grantees with small numbers of classrooms) these tests have low power to detect statistically significant differences.

Multiple Comparisons

No adjustments were made to the *p*-values from the results of this study before comparing them to the significance levels, indicated in the report by asterisks. Multiple outcomes were examined with the same children, teachers, and classrooms, and within each outcome multiple time points were examined and multiple contrasts tested. Therefore, the overall "family-wise" type I error rate is not protected at a particular significance level by only considering results significant at that level or lower. If a conservative line aimed at protecting against type I errors at all costs were adopted, the power to detect effects would decline. Arguments have been made by some that such adjustments are not necessary for a variety of reasons (Feise 2002; Rothman 1990; Savitz and Olshan 1995). Among these reasons is the suggestion that in some studies power considerations are more important than type I errors. The nature of this evaluation was more exploratory, and sought significant effects for curricula measured against an already high bar found in most prevailing evaluations. The reader may not wish to put too much weight on the *p*-values, although significance levels are reported for the results. The possibility of spurious results are possible, but some protection against them was afforded by the decision not to accept a curriculum as having made a significant difference in a domain unless it was found to do so on more than one measure in that domain. Also, the focus on effect sizes provides a better sense of the amount of improvement afforded by the various intervention curricula on each measure.

Effect Size Calculation

Definition of Effect Size

The significant intervention effects are determined by tests of linear combinations of beta coefficients estimated for the model; the reflection of that difference is reported as an effect size in the main report and in appendix A. Effect sizes are often used to provide a relative measure of the magnitude of differences. In this report, the effect size is defined as the difference between two statistics of interest divided by a normalizing factor:

$$d = \frac{\delta\mu}{\sigma} = \frac{\mu_2 - \mu_1}{\sigma} \qquad (10)$$

The statistics of interest, μ_i 's, could be means of experimental groups at a particular time point (such as the mean of an intervention group at fall pre-kindergarten), differences in group means at two different time points, as well as slopes of the growth curves. Both the numerator (the difference) and denominator (the normalizing factor) need to be estimated from the sample data.

The function of the normalizing factor is to remove the influence of population variation from the measure of interest. Many quantities can be used for this factor: the estimated standard deviation (SD) of the control group at baseline or an average of the estimated standard deviations of the two groups at baseline. When the groups are created from the same population using proper randomization procedures, the underlying population variation for them is the same and pooling estimates from the two groups increases the reliability of the estimated SD.

Calculation of the Numerator

The numerators for the effect size calculations are differences of means of control and intervention groups at various time points and differences in slopes of the growth curves of different groups between various time points. Nevertheless, they can all be expressed in terms of linear combinations of parameters of the models we fit, either repeated measures or ANCOVA models. As such, the general formula for estimating an effect size is given as

$$\hat{d} = \frac{C'\beta}{\hat{\sigma}_{pooled}} = \frac{\beta_2 - \beta_1}{\hat{\sigma}_{pooled}} \tag{11}$$

where β's are the model parameters and σ_{pooled} the pooled estimated population standard deviation as described above.

For comparisons of means, either child or classroom outcomes, the difference in the two betas shown in (11) becomes the difference between two means. For example, for the spring pre-kindergarten comparison of control vs. intervention, using $\hat{M}_{Pre-K,S,C}$ as the estimated mean for the control group at the spring pre-kindergarten assessment and $\hat{M}_{Pre-K,S,T}$ the estimated mean for the intervention group, and σ_{pooled} the appropriate pooled estimated standard deviation for the outcome of interest (either child or classroom), is defined as:

$$\hat{d}_{Pre-K,S,means} = \frac{\hat{M}_{Pre-K,S,T} - \hat{M}_{Pre-K,S,C}}{\hat{\sigma}_{pooled}} \tag{12}$$

Similarly, for child or classroom outcomes in slope differences, Cohen's d for the pre-kindergarten slope comparison of control vs. intervention at the Pre-k, spring time point can be defined as follows. Let $\beta_{pre-k,S,C}$ be the pre-kindergarten slope for the control population at the spring data collection, $\beta_{Pre-K,S,T}$ be the same for the intervention population, and again σ_{pooled} is the pooled estimated standard deviation for whichever child or classroom outcome is of interest. The effect size for the difference in slopes for the pre-kindergarten year would be:

$$\hat{d}_{Pre-K,S,Slopes} = \frac{\beta_{Pre-K,S,T} - \beta_{Pre-K,S,C}}{\hat{\sigma}_{pooled}} \tag{13}$$

Choice of the Normalizing Factor

For the PCER study data, the statistics of interest for which effect sizes are calculated include differences between the means of control and intervention groups at various time points, and differences between the slopes of the growth curves of control and intervention groups between various time points. As such, the different statistics of interest are estimated, using data from different time points. In determining the most appropriate normalizing factor for the calculation of the effect sizes, besides requiring that it reflect the variation of the population under study, a common approach for all outcomes and, for each outcome, a common normalizing factor that can be used for all appropriate statistics and for as many sites as possible (except sites with different designs and outcomes with fewer time points) was sought. This reduced the complexity of the calculation and made it easier to understand, but also provides a unified platform on which results are compared. In addition, using a uniform approach made significant results less likely due to happenstance.

One possible argument against this approach is that since the knowledge and skills of the young children change over time, so does the reference population. Thus, it is necessary to investigate change in the population variation. For this purpose, the population variation was estimated by time point for each of the outcomes. After averaging the estimated population standard deviations across group assignment and site, the results are summarized in table B-6.

Table B-6. Average estimated population standard deviation by study outcome and time points

Outcome	Fall Pre-K	Spring Pre-K	Spring K	Average
PPVT	16.56	16.76	15.09	16.13
Pre-CTOPPP	3.64	4.12	—	3.88
CMA-A Mathematics Composite	0.23	0.23	0.17	0.21
TOLD	4.77	4.64	3.86	4.42
Shape Composition[1]	0.85	0.87	0.80	0.84
WJ Letter Word Identification	25.09	23.96	28.61	25.89
WJ Applied Problems	21.51	18.42	17.24	19.06
WJ Spelling	25.93	25.39	22.43	24.58
TERA	6.88	8.98	9.81	8.56
SSRS Social Skills	14.87	14.17	—	14.52
SSRS Problem Behaviors	13.02	12.82	—	12.92
PLBS	10.08	10.26	—	10.17
Average	11.94	11.72	12.25	11.94

— Not available.

[1] Building Blocks, Shape Composition task

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

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

As can be seen, for most outcomes, only small changes occur in population variation between the first two time points (fall pre-kindergarten and spring pre-kindergarten). There are indications, however, that the population variation becomes smaller at the third time point (spring kindergarten). One possible explanation is that children perform more uniformly once they are in a structured school setting such as kindergarten. The results also suggest that population variation of the first two times should be considered as the benchmark for population variation. To increase the reliability of the sample estimates, the weighted average of the estimated population variation across group assignment and the first two time points as the normalizing factor in the calculation of the effect sizes was used.

To assess the impact of pooling the estimated population variation between the control and intervention groups, the variance estimates between the two groups were investigated. Averaging across site, and all available time points, the results are summarized in table B-7.

Again, for most of the outcomes, when averaging over site and time point, the difference between estimated population standard deviation of the control and intervention groups was minimal.

Table B-7. Average estimated population standard deviation by study outcome and group assignment

Outcome	Control group	Intervention group	Average
PPVT	16.78	15.59	16.13
Pre-CTOPPP	3.92	3.85	3.88
CMA-A Mathematics Composite	0.22	0.21	0.21
TOLD	4.57	4.30	4.42
Shape Composition[1]	0.86	0.82	0.84
WJ Letter Word Identification	26.30	25.53	25.89
WJ Applied Problems	20.31	17.98	19.06
WJ Spelling	24.70	24.48	24.58
TERA	8.72	8.41	8.56
SSRS Social Skills	14.32	14.69	14.52
SSRS Problem Behaviors	12.96	12.89	12.92
PLBS	10.36	10.01	10.17
Average	12.20	11.71	11.94

[1] Building Blocks, Shape Composition task
NOTE: Refer to the glossary for abbreviations of the measures. When there are two intervention groups, the number for intervention is the average of the two standard deviations.
SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Calculation of the Normalizing Factor

The analyses focused on a single outcome measure at a time. Modeling and estimation were carried out site by site for each intervention curriculum. Thus, for each measure and site combination, a normalizing factor needed to be obtained, requiring two steps: (1) determine the appropriate data points to be included for the calculation, and (2) the actual calculation of the pooled standard deviation. Table B-8 provides an example of how the data were selected for this calculation under a typical situation: a grantee with a single site and a single intervention where data from three time points are available. Cells used in estimating the pooled population standard variation are indicated as "used in pooling". Others are not used in the calculation.

Table B-8. Pooled standard deviation example

Intervention or control	Fall Pre-K	Spring Pre-K	Spring K
Intervention	Used in pooling	Used in pooling	Not used in pooling
Control	Used in pooling	Used in pooling	Not used in pooling

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

This could be data on a Woodcock-Johnson measure from a single site grantee such as New Hampshire. The variances and sample sizes for the data of the four cells (intervention and control groups for fall pre-kindergarten and spring pre-kindergarten assessment data) are pooled, using a weighted combination (see below) to arrive at the pooled standard deviation. The pooled standard deviation obtained with these data is used to calculate effect size estimates for comparisons made between control and intervention groups for the longitudinal spline model as well as the two ANCOVA models of spring pre-kindergarten and kindergarten outcomes. There are many exceptions to the typical example illustrated in table B-8. For example, there are sites where more than one intervention is used. There are also outcome measures where data from only two time points are available. The section entitled Pooled Standard Deviation Calculation Details for Each Data Structure by Research Team Site/Intervention Configuration, provides further details on the data selection step for grantees with more complicated designs and outcomes for which we collected data at different time points.

Once the data points used for pooling are determined, calculation of the pooled standard deviation is carried out by fitting a null model (with no covariates) with a classroom-level random effect. For example, for each cell in table B-8 above, a model as follows is fit:

$$y_{i,j} = \mu_0 + u_i + \varepsilon_{i,j} \tag{14}$$

where i, j are indices for classroom and children within a classroom, respectively; $y_{i,j}$ is the outcome under study; μ_0 the intercept term; and u_i and $\varepsilon_{i,j}$ are the classroom random effect and individual residual error, respectively. The resulting estimated variance components from fitting model (14) are the classroom variance component, $\hat{\sigma}_u^2$ associated with u_i ‘s and the residual variance, $\hat{\sigma}_\varepsilon^2$, associated with $\varepsilon_{i,j}$ ‘s. The sum of the two components provides the total population variation for the outcome/site/intervention/time point combination.

$$\hat{\sigma}^2_{total,trt,k} = \hat{\sigma}^2_{u,trt,k} + \hat{\sigma}^2_{\varepsilon,trt,k} \tag{15}$$

where trt and tp are indicators for intervention group assignment and time point, respectively.

Next, the pooled population variance estimate is calculated as a weighted average of estimates (15) from all cells as

$$\hat{\sigma}^2_{Pooled} = \left(\sum\nolimits_{trt,k}^{K} (n_{trt,tp} - 1)\hat{\sigma}^2_{Total,trt,k}\right) \Big/ \left(\sum\nolimits_{trt,k}^{K} (n_{trt,k} - 1)\right) \tag{16}$$

where K is the total number of cells created by the intervention by time point combination for the particular outcome and site. Because the sample sizes within each cell are different, the weighting by relative sample sizes produces an unbiased estimate of the population variance.

The method of estimating population variance used here can be validated by examining the sample variance without considering the hierarchical nature of the sample. Table B-9 shows the population standard deviation for twelve child outcomes as estimated using SAS PROC MEANS compared with estimates using the mixed model (14), as described above. The results are remarkably close. Since estimates from SAS PROC MEANS do not take into account the hierarchical structure of the sample data, those estimates were consistently slightly smaller than estimates from the mixed model, validating the additional variance components that have been accounted for by the results from the mixed model (14).

Table B-9. Estimated pooled population standard deviation using unconditional standard deviations and standard deviations from repeated measures analyses

Outcome	Mixed model	SAS PROC MEANS
PPVT	17.3826	17.2059
Pre-CTOPPP	3.9495	3.9210
CMA-A Mathematics Composite	0.2377	0.2355
TOLD	4.8631	4.8343
Shape Composition[1]	0.8948	0.8887
WJ Letter Word Identification	25.4034	25.0996
WJ Applied Problems	21.5521	21.3774
WJ Spelling	25.9873	25.6834
TERA	8.2939	8.2063
SSRS Social Skills	15.0638	14.8298
SSRS Problem Behaviors	13.3030	13.1097
PLBS	10.5641	10.4620

[1] Building Blocks, Shape Composition task

NOTE: The standard deviations were calculated using SAS PROC MEANS and Mixed Model averaged over sites. Refer to the glossary for abbreviations of the measures.

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

Pooled Standard Deviation Calculation Details for Each Data Structure by Research Team Site/Intervention Configuration

This section provides additional details on data pooling for the calculation of the normalization factor used in reporting the effect sizes. This study features a variety of designs (four in all) among the PCER study grantees. The tables (table B-10 through table B-12) generalize table B-8 to cover the various situations encountered in our analysis, both in terms of design and the time points at which data on an outcome were collected. For each of our team site/intervention configurations, the tables contain a row for each intervention group within the given configuration. Within these rows, the cells (group by time period combinations) of data used in estimating a given pooled population standard variation for a particular site or intervention within a team are indicated by the inclusion of the same individual letter (A, B, or C) in that set of cells. "A" for site 1 or intervention 1, "B" for site 2 or intervention 2, et cetera. Teams with two or three sites had results reported at the individual site level and combined across the grantees sites. Pooling for such combined results for double or triple site grantees are indicated by letter combinations (e.g., AB or ABC). Those cells not used in any calculation for that set of outcomes and analyses are marked with an "x." Cells for which no such data exist are marked with a "-." Thus, for example, for a child outcome with three time points of data at a single site/double intervention (such as Vanderbilt University, University of Texas-Houston, and Florida State University), the "A"s in four of table B-10's cells of the appropriate intervention 1 and control rows under the "Intervention 1 Test" columns indicate that these four cells were used in the calculating the pooled standard deviation used for estimating the effect size for intervention 1. Similarly, the four cells that contain "B"s in the rows for intervention 2 and control are pooled to form the pooled standard deviation for estimating effect sizes for intervention 2. Another example is the case of a research team working at two sites with a single intervention (i.e., the University of California, Berkeley with the University at Buffalo, State University of New York (California/New York) working in both California and New York) with data at all three time points. The "A"s could indicate pooling for the California site's pooled standard deviation used in calculating their effect sizes, the "B"s pooling for the New York site's pooled standard deviation and effect size calculations, and the "AB"s indicate cells involved in pooling for combined California/New York pooled standard deviations used in effect size estimations. Table B-10 contains the pooling used for the bulk of the analysis, the spline models, repeated measures with pre-kindergarten data, and pre-kindergarten and spring kindergarten ANCOVAs for which pre-kindergarten measures were the same as kindergarten measures. Table B-11 contains the pooling for the ANCOVAs with TBRS outcomes. Table B-12 contains the pooling for spring kindergarten ANCOVAs on measures that changed between pre-kindergarten and kindergarten assessments.

Table B-10. Pooled standard deviation details: Pooling for outcomes modeled with the simple repeated measures, the repeated measures spline models, the pre-kindergarten spring analysis of covariance (ANCOVA) models (except Teacher Behavior Rating Scale (TBRS)), and the kindergarten spring ANCOVA models where kindergarten data were comparable to pre-kindergarten

Intervention and control group combinations	Fall Pre-K	Spring Pre-K	Spring K
Single site/single intervention			
Intervention	A	A	X
Control	A	A	X

Intervention and control group combinations	Fall Pre-K		Spring Pre-K		Spring K
	Intervention 1 test	Intervention 2 test	Intervention 1 test	Intervention 2 test	
Single site/double intervention					
Intervention 1 Sites	A	†	A	†	X
Intervention 2 Sites	†	B	†	B	X
Control	A	B	A	B	X

Intervention and control group combinations	Fall Pre-K		Spring Pre-K		Spring K
	Single site tests	Combined site tests	Single site tests	Combined site tests	
Double site/single intervention					
Intervention group-Site 1	A	AB	A	AB	X
Control group-Site 1	A	AB	A	AB	X
Intervention group-Site 2	B	AB	B	AB	X
Control group-Site 2	B	AB	B	AB	X

Intervention and control group combinations	Fall Pre-K		Spring Pre-K		Spring K
	Single site tests	Combined site tests	Single site tests	Combined site tests	
Triple site/single intervention					
Intervention group-Site 1	A	ABC	A	ABC	X
Control group-Site 1	A	ABC	A	ABC	X
Intervention group-Site 2	B	ABC	B	ABC	X
Control group-Site 2	B	ABC	B	ABC	X
Intervention group-Site 3	C	ABC	C	ABC	X
Control group-Site 3	C	ABC	C	ABC	X

† Not applicable.

NOTE: A: Intervention group-Site 1, Control group-Site 1
B: Intervention group-Site 2, Control group-Site 2
C: Intervention group-Site 3, Control group-Site 3
X: Data were not used in the calculations.

Details about reading this table are found in the section Pooled Standard Deviation Calculation Details for Each Data Structure by Team Site/Intervention Configuration. A simpler example is illustrated in table B-8. The repeated measures spline model was used to analyze data collected at three time points (fall and spring of pre-kindergarten and spring of kindergarten). The simple repeated measures model was used to analyze data collected at two time points (fall and spring of pre-kindergarten).

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

Table B-11. Pooled standard deviation details: Pooling for Teacher Behavior Rating Scale (TBRS) outcomes modeled with the pre-kindergarten spring analysis of covariance (ANCOVA) models

Intervention and control group combination	Fall Pre-K	Spring Pre-K	Spring K
Single site/single intervention			
Intervention	—	A	—
Control	—	A	—

Intervention and control group combination	Fall Pre-K	Spring Pre-K: Intervention 1 test	Spring Pre-K: Intervention 2 test	Spring K
Single site/double intervention				
Intervention 1	—	A	†	—
Intervention 2	—	†	B	—
Control	—	A	B	—

Intervention and control group combination	Fall Pre-K	Spring Pre-K: Single site tests	Spring Pre-K: Combined site tests	Spring K
Double site/single intervention				
Intervention group-Site 1	—	A	AB	—
Control group-Site 1	—	A	AB	—
Intervention group-Site 2	—	A	AB	—
Control group-Site 2	—	A	AB	—

Intervention and control group combination	Fall Pre-K	Spring Pre-K: Single site tests	Spring Pre-K: Combined site tests	Spring K
Triple site/single intervention				
Intervention group-Site 1	—	A	ABC	—
Control group-Site 1	—	A	ABC	—
Intervention group-Site 2	—	B	ABC	—
Control group-Site 2	—	B	ABC	—
Intervention group-Site 3	—	C	ABC	—
Control group-Site 3	—	C	ABC	—

— Not available.
† Not applicable.
NOTE: A: Intervention group-Site 1, Control group-Site 1
B: Intervention group-Site 2, Control group-Site 2
C: Intervention group-Site 3, Control group-Site 3.
Details about reading this table are found in the section Pooled Standard Deviation Calculation Details for Each Data Structure by Team Site/intervention Configuration. A simpler example is illustrated in table B-8.
SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Table B-12. Pooled standard deviation details: Pooling for kindergarten spring outcomes (SSRS, Pre-CTOPPP/CTOPP, PLBS/LBS) modeled with analysis of covariance (ANCOVA) models

Intervention and control group combination	Fall Pre-K	Spring Pre-K	Spring K	
Single site/single intervention				
Intervention	—	—	A	
Control	—	—	A	
Intervention and control group combination	**Fall Pre-K**	**Spring Pre-K**	**Spring K: Intervention 1 test**	**Spring K: Intervention 2 test**
Single site/double intervention				
Intervention 1	—	—	A	†
Intervention 2	—	—	†	B
Control	—	—	A	B
Intervention and control group combination	**Fall Pre-K**	**Spring Pre-K**	**Spring K: Single site tests**	**Spring K: Combined site tests**
Double site/single intervention				
Intervention group-Site 1	—	—	A	AB
Control group-Site 1	—	—	A	AB
Intervention group-Site 2	—	—	B	AB
Control group-Site 2	—	—	B	AB
Intervention and control group combination	**Fall Pre-K**	**Spring Pre-K**	**Spring K: Single site tests**	**Spring K: Combined site tests**
Triple site/single intervention				
Intervention group-Site 1	—	—	A	ABC
Control group-Site 1	—	—	A	ABC
Intervention group-Site 2	—	—	B	ABC
Control group-Site 2	—	—	B	ABC
Intervention group-Site 3	—	—	C	ABC
Control group-Site 3	—	—	C	ABC

— Not available.
† Not applicable.
NOTE: A: Intervention group-Site 1, Control group-Site 1
B: Intervention group-Site 2, Control group-Site 2
C: Intervention group-Site 3, Control group-Site 3
Details about reading this table are found in the section Pooled Standard Deviation Calculation Details for Each Data Structure by Team Site/intervention Configuration. A simpler example is illustrated in table B-8. Refer to the glossary for abbreviations of the measures.
SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Model Specification

Standardization of Covariates

To enhance interpretability of model estimates, all continuous covariates were standardized (mean = 0, standard deviation = 1). Two covariates were standardized in the child outcome models: the child's age and the fall baseline measure. Thus, child outcome model estimates are for a child at the mean value of all continuous covariates and an average of the effects for all categories for each categorical covariates (e.g., average age, average fall baseline, averages across effects for child's gender, race/ethnicity, maternal education and parent reported individual education plan categories). For the classroom outcome models, the four continuous covariates that were standardized include previous years of preschool teaching, child-to-adult ratio, average class size, and the fall baseline measure. The standardization of these covariates means the classroom outcome estimates are specified for a classroom considered to have an average child to adult ratio, class size, and fall baseline score, and taught by a teacher with an average amount of preschool teaching experience and equally averaged across the categories of the classification covariates: city size, the teacher's BA attainment, and teacher's race.

Data Clustering—Nesting of Children in Kindergarten Classrooms and Blocking

Analysis of educational data such as these in which children reside in fixed classrooms must take into account the common experience they share that tends to reduce their individual variance. The data clustering introduced by the use of intact classrooms can be included by introducing such nesting into the model. As indicated above, our model incorporated this by estimating random classroom intercepts. In these multi-site longitudinal hierarchical data there were two possible nesting structures (represented by the random classroom intercepts) that could have been used in the analyses: the pre-kindergarten and kindergarten classroom nestings. The nesting of children in the pre-kindergarten classrooms was by study design (i.e., multiple children typically were selected from pre-kindergarten study classrooms), whereas nesting in the kindergarten classrooms reflected the extent to which children in the pre-kindergarten programs attended the same or different primary schools. In addition, intervention was implemented in the pre-kindergarten classrooms. Therefore, it was logical to use the pre-kindergarten classrooms as the nesting structure in our models.

We examined the clustering of the data in the pre-kindergarten and kindergarten years separately by intervention group and site. Table B-13 displays the range of cluster sizes for our data in the pre-kindergarten- and kindergarten-year classrooms. Though not provided in the table, a general description of the distributions of these sites by intervention-group cluster sizes is provided below. The distributions in the kindergarten year are all highly right skewed with the majority of the classrooms having a frequency of one child per class, and much lower frequencies for any larger class clusters. The data in the pre-kindergarten year are more variable, though practically never right skewed, and then only slightly. The data distribution is generally flat or unimodal, with a few being left skewed (lower frequencies for smaller class sizes and higher frequencies for larger class sizes). In most cases the pre-kindergarten minimum cluster size is larger than the kindergarten maximum cluster size, indicating the relative rareness of kindergarten year clustering. In virtually every case of overlap in the cluster size distributions (i.e., the maximum kindergarten class size is larger than the minimum pre-kindergarten class size), the overlap results from a single classroom, which tends to be an individual outlier among the rest of the classes within that site by intervention group in that year.[3] Fully 37

[3] For example, the site 4 control group shows some overlap, but only one pre-kindergarten classroom had only 5 children and the next smallest classroom had 10 children. Only one classroom in kindergarten had six and the next largest had only four. Site 15 has the most overlap. In the pre-kindergarten year the control group had one classroom with one student and one classroom with two. All the others had four or more. In the kindergarten year, the site had one classroom each with five, four, and three children, and all the others are two or less. Similarly in the intervention group, in the pre-kindergarten year only one classroom each had two and four children clustered in it, while in the kindergarten year only one each had eight, six, or four children clustered in the same classrooms.

percent of the kindergarten clusters were singletons (having only one PCER child in the classroom) and another 20 percent were clusters of two. In pre-kindergarten 95 percent of the clusters were larger than five while in kindergarten only 9 percent were this large. It seemed clear that the pre-kindergarten clustering was preferred for this analysis, as cluster sizes in kindergarten are so small, especially relative to the pre-kindergarten clusters, and have only a minimal impact on the estimation of error variability. In fact, within each team/site, very often multiple preschools are in the study. However, since we consider the classroom to be the most influential clustering structure, it is the only structure reflected in the model.

Table B-13. Pre-kindergarten and kindergarten classroom clusters of children, maximum and minimum size

		Pre-kindergarten		Kindergarten	
Site	Group	Maximum	Minimum	Maximum	Minimum
1	Control	9	7	5	1
1	Intervention	8	8	4	1
2	Control	9	6	6	1
2	Intervention	8	7	4	1
3	Control	10	4	4	1
3	Intervention	14	7	6	1
4	Control	16	5	6	1
4	Intervention	14	3	4	1
5	Control	7	5	2	1
5	Intervention	10	8	4	1
6	Control	11	6	7	1
6	Intervention	16	4	4	1
7	Control	14	11	3	1
7	Intervention	13	12	4	1
8	Control	11	8	3	1
8	Intervention	11	8	2	1
9	Control	19	8	7	1
9	Intervention	19	10	8	1
10	Control	7	3	5	1
10	Intervention	9	5	5	1
11	Control	17	10	12	1
11	Intervention	17	16	11	1
12	Control	21	5	7	1
12	Intervention	20	2	5	1
13	Control	14	7	8	1
13	Intervention	15	5	6	1
14	Control	8	5	4	1
14	Intervention	8	5	4	1
15	Control	8	1	5	1
15	Intervention	10	2	8	1
16	Control	11	6	2	1
16	Intervention	15	3	3	1
17	Control	13	12	10	1
17	Intervention	13	8	4	1
18	Control	17	9	11	1
18	Intervention	17	12	7	1

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

We also should note that there is also site-level blocking that occurs in the site data, as randomization takes place within sites. The team-level analysis that we conducted takes this into account with site-specific effect coding whenever multiple sites were involved for a given team.

Covariance Structure and Modeling Steps

To account for the clustering mentioned above as well as the correlated data for individual children or classrooms across time points, a mixed model was used which features means of accounting for these sources of covariance. The covariance structure for each of the models was determined in the initial steps of conducting those analyses, one for random effects in the model accounting for the classroom clustering and the other for the correlations among the repeated measures.

In determining the covariance structure of the random effects, there were three random effects were considered in the linear spline model (intercept, pre-kindergarten slope, and kindergarten slope), two in the simple repeated measures model (intercept and pre-kindergarten slope), and a single random intercept in the ANCOVA models of child outcomes. Preliminary model fitting looked at both a variety of structures, beginning with a completely unstructured covariance that included random effect variances for the intercept, pre-kindergarten slope, and kindergarten slope and their correlations. Whether there was sufficient randomness found to justify including each of these effects in the model as random was checked. These latter checks were done using nested likelihood tests with critical points from the appropriate 50:50 mixture of chi-squared distributions (Fitzmaurice, Laird, and Ware 2004). The former were examined by considering comparative values for appropriate information criteria (Akaike's Information Criterion [AIC], Bayesian Information Criterion [BIC], and the corrected Akaike's Information Criterion [AAIC]). In the vast majority of cases, the variance estimate for the classroom intercept random effect was significant, indicating that if not accounted for, the total variation would be underestimated. Random pre-kindergarten and kindergarten slopes were generally not statistically significant and, except for the repeated measures model for the child behavioral outcomes, were not included as random effects in the final model. The random classroom intercept was included in all repeated measures and ANCOVA models for individual child outcomes. For classroom outcome models classroom clustering is not an issue since classrooms are the level of analysis, thus no random effects were included in those models.

The most appropriate covariance structures for the residual variances across the time points were identified using nested model comparisons when possible and information criteria when nesting was not possible. This was done by starting with unstructured models then examining a wide variety of structures that appeared possible, based on the estimates from the unstructured models. In general, we found that several covariance structures fit the data equally well (i.e., the data did not strongly indicate that one structure was much better than any other). Resulting covariance structures used for each final model are detailed in table B-14 and their definitions are given in table B-15.

The covariance structure was chosen and the mean model structure refined, where necessary, using nested model comparisons with maximum likelihood (ML) estimation. Finally, variance components were estimated using REML for the final model estimation, providing unbiased results for testing of group differences and other results tabled in this report. A more detailed discussion concerning REML versus ML estimation was given in the section on Estimation. These analyses were conducted at the research team level, with separate estimates for sites and curricula within sites for each team.

Table B-14. Specific covariance structures found to best fit the data

Variables analyzed	Analysis model	Random variables and any covariance structure (G)	Repeated Measures residual covariance structure (R)
		Analyses reported in the main report[1]	
Eight child cognitive outcomes/ three time points	Repeated Measures Spline Model	Classroom intercept	Heterogeneous compound symmetry
Cognitive outcome/two time points	Simple Repeated Measures	Classroom intercept	Unstructured
Behavioral outcomes/two time points	Simple Repeated Measures	Classroom intercept and time/variance components	Unstructured
Classroom outcomes/one time point	ANCOVA	None	Simple residual variance
		Analyses reported in appendix A	
Eight child cognitive outcomes/ thee time points	Repeated Measures Spline Model	Classroom intercept	Heterogeneous compound symmetry
Cognitive outcome/two time points	Simple Repeated Measures	Classroom intercept	Unstructured
Behavioral outcomes/two time points	Simple Repeated Measures	Classroom intercept and time/variance components	Unstructured
Eight child cognitive outcomes/ three time points	ANCOVA	Classroom intercept	Simple residual variance
Cognitive outcome/two time points	ANCOVA	Classroom intercept	Simple residual variance
Behavioral outcomes/two time points	ANCOVA	Classroom intercept	Simple residual variance
Classroom outcomes/two time points	ANCOVA	None	Simple residual variance

[1] The term main report refers to chapters 1-13.
NOTE: ANCOVA: Analysis of covariance. The repeated measures spline model was used to analyze data collected at three time points (fall and spring of pre-kindergarten and spring of kindergarten). The simple repeated measures model was used to analyze data collected at two time points (fall and spring of pre-kindergarten).
SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Table B-15. Specific covariance structure definitions

Covariance structure	Definition
Heterogeneous compound symmetry	The *(i,j)*th element of the covariance matrix is defined as[1]: $\sigma_i \sigma_j(\rho 1(i \neq j) + 1(i = j))$
Variance components	The *(i,j)*th element of the covariance matrix is defined as: $\sigma^2_k 1(i = j)$ and *i* corresponds to the *k*th effect. This means the variances for the random intercepts and random slopes were each freely estimated, and the covariance between them is fixed at zero.
Unstructured	The *(i,j)*th element of the covariance matrix is defined as: σ_{ij} This structure allows the model to freely estimate the random variance for the two time points, as well as the covariance between them.

[1] The formalization, $1(i \neq j)$, assigns 1 to this term if $i \neq j$; the formalization $1(i = j)$, assigns 1 to the term if $i = j$.
SOURCE: The Preschool Curriculum Evaluation Research (PCER) Study.

Test for Equality of Pre-kindergarten and Kindergarten Period Slopes

The linear spline model estimates separate intervention group growth trajectories for the pre-kindergarten instructional period (projected back to the start of the intervention and tested for differences in the tables) and for the period following the pre-kindergarten instructional period, ending in the spring of the kindergarten year. If the rate of change during the two time periods is the same, then this model does not need to include a separate slope term for the kindergarten period. To check for such a possible model simplification, equality of slopes for the preschool and kindergarten years was tested. This simplification did not fit the data well; indicating rates of growth did in fact differ across the pre-kindergarten and kindergarten periods.

Site by Covariate Interactions

Since all modeling was done at the research team and curriculum level, site by covariate interactions are ruled out in most cases. These were checked for teams with multiple sites. The vast majority of these were nonsignificant, and thus these terms were not included in the models.

Homogeneity of Regression Assumption

As part of our preliminary model checking, the equality of the covariates' slopes were checked for equality across intervention groups by including intervention by covariate interaction terms. Table B-16 contains the significant interactions that were found. While this number of significant interactions may seem large, this is only 10 percent of the 840 possible interactions. Only one factor, disability status, emerged somewhat consistently, accounting for 23 percent of the "significant" interactions. However, the very small number of children with disabilities in many intervention or control groups resulted in rather odd and unstable interactions. Therefore, all intervention x covariate interactions in subsequent analyses were excluded.

Group Comparisons[4] Testing for Intervention Impact

All intervention impact comparisons were conducted at the curriculum level, because each research team selected and implemented its intervention independently. Accordingly, intervention differences in means and rates of change over time had to be tested at the team level. This means all data for a given team's sites were analyzed in the same model, with terms included to separate out effects for different curricula or sites for teams with two intervention curricula, or two geographic locations and one intervention curriculum. As stated previously, the repeated measures models allowed a simultaneous test of whether intervention differences in the spring score(s), and the rates of change between specified time points (e.g., fall-spring pre-kindergarten or spring pre-kindergarten to kindergarten). These means and rates of change were estimated across sites for grantees with multiple geographic locations (e.g., the combined effect of *Pre-K Mathematics with DLM Early Childhood Express Math software* package in California and New York). Statistical tests for mean or slope intervention impacts were conducted using t-tests with estimated degrees of freedom, as explained below. All tests were two-tailed tests and the significance level was indicated with asterisks at .05, .01, .001, and .0001 significance levels (see tables with the impact analysis results in the main report and appendix A).

[4] Group comparisons refer to treatment versus control group comparisons for each intervention.

Table B-16. Significant treatment by covariate interactions from check on homogeneity of regression assumption

Variable	Effect	Prob F
TERA	ELLM_BAY * Child'sAge	0.0306
TERA	ELLM_BAY * DisabilityProxyFall	0.0028
TERA	ELLM_JCK * Child'sGender	0.0258
TERA	CC(UNC)_NC * DisabilityProxyFall	0.0296
TERA	BB_TN * Child'sAge	0.0164
TERA	PA_WI * Child'sAge	0.0430
TERA	PC_MO * DisabilityProxyFall	0.0360
TERA	LE_FL * Child'sAge	0.0249
TERA	DLM with OC_FL * Child'sRace	0.0161
TERA	Curiosity Corner_FL * MaternalEducation	0.0258
TERA	Curiosity Corner_NJ * DisabilityProxyFall	0.0029
WJ Applied Problems	ELLM_JCK * Child'sGender	0.0182
WJ Applied Problems	ELLM_MIA * Child'sAge	0.0340
WJ Applied Problems	CC (UNC)_NC * MaternalEducation	0.0318
WJ Applied Problems	CC(UNC)_NC * DisabilityProxyFall	0.0019
WJ Applied Problems	LB_TX * DisabilityProxyFall	0.0137
WJ Applied Problems	PC_MO * DisabilityProxyFall	0.0008
WJ Applied Problems	LE_FL * DisabilityProxyFall	0.0417
WJ Applied Problems	Curiosity Corner_KS * Child'sAge	0.0497
WJ Applied Problems	Curiosity Corner_KS * DisabilityProxyFall	0.0198
WJ Applied Problems	Curiosity Corner_NJ * MaternalEducation	0.0147
WJ Applied Problems	Curiosity Corner_NJ * DisabilityProxyFall	0.0022
Shape Composition[1]	ELLM_JCK * Child'sAge	0.0454
Shape Composition[1]	CC(UNC)_NC * DisabilityProxyFall	0.0106
Shape Composition[1]	LE_FL * DisabilityProxyFall	0.0033
Shape Composition[1]	Curiosity Corner_FL * Child'sGender	0.0108
Shape Composition[1]	Curiosity Corner_FL * Child'sAge	0.0018
Shape Composition[1]	Curiosity Corner_NJ * MaternalEducation	0.0030
Shape Composition[1]	LFC_VA * Child'sGender	0.0204
WJ Letter Word Identification	Pre-K Math_CA * Child'sAge	0.0087
WJ Letter Word Identification	Pre-K Math_NY * MaternalEducation	0.0331
WJ Letter Word Identification	ELLM_BAY * DisabilityProxyFall	0.0369
WJ Letter Word Identification	ELLM_MIA * DisabilityProxyFall	0.0036
WJ Letter Word Identification	CC(UNC)_NC * DisabilityProxyFall	< 0.0001
WJ Letter Word Identification	BB_TN * Child'sAge	0.0303
WJ Letter Word Identification	DD_TX * DisabilityProxyFall	0.0470
WJ Letter Word Identification	PC_MO * DisabilityProxyFall	0.0079
WJ Letter Word Identification	Curiosity Corner_KS * DisabilityProxyFall	0.0008
WJ Letter Word Identification	Curiosity Corner_FL * MaternalEducation	0.0124
WJ Letter Word Identification	Curiosity Corner_NJ * MaternalEducation	0.0205
WJ Letter Word Identification	Curiosity Corner_NJ * DisabilityProxyFall	0.0156

See notes at end of table.

Table B-16. Significant treatment by covariate interactions from check on homogeneity of regression assumption—Continued

Variable	Effect	Prob F
WJ Spelling	Pre-K Math_CA * DisabilityProxyFall	0.0274
WJ Spelling	ELLM_BAY * DisabilityProxyFall	0.0044
WJ Spelling	CC(UNC)_NC * DisabilityProxyFall	0.0102
WJ Spelling	CC(UNC)_TN * Child'sRace	0.0310
WJ Spelling	CC(UNC)TN * DisabilityProxyFall	0.0147
WJ Spelling	BB_TN * Child'sRace	0.0261
WJ Spelling	LB_TX * DisabilityProxyFall	0.0008
WJ Spelling	DD_TX * DisabilityProxyFall	0.0047
WJ Spelling	PC_MO * Child'sGender	0.0476
WJ Spelling	PC_MO * DisabilityProxyFall	0.0066
WJ Spelling	LE_FL * Child'sAge	0.0006
WJ Spelling	LE_FL * DisabilityProxyFall	0.0416
WJ Spelling	Curiosity Corner_KS * DisabilityProxyFall	< 0.0001
WJ Spelling	Curiosity Corner_FL * Child'sGender	0.0082
WJ Spelling	Curiosity Corner_FL * MaternalEducation	0.0332
WJ Spelling	Curiosity Corner_NJ * MaternalEducation	0.0497
PPVT	ELLM_MIA * Child'sAge	0.0060
PPVT	ELLM_MIA * Child'sRace	0.0411
PPVT	CC(UNC)_NC * DisabilityProxyFall	0.0011
PPVT	CC(UNC)_TN * Child'sGender	0.0376
PPVT	PA_WI * Child'sRace	0.0122
PPVT	Curiosity Corner_KS * Child'sGender	0.0162
PPVT	Curiosity Corner_FL * MaternalEducation	0.0291
TOLD	Pre-K Math_CA * Child'sAge	0.0305
TOLD	ELLM_MIA * Child'sAge	0.0054
TOLD	CC(UNC)_NC * DisabilityProxyFall	0.0414
TOLD	LB_TX * DisabilityProxyFall	0.0177
TOLD	DD_TX * DisabilityProxyFall	0.0348
TOLD	PA_WI * Child'sGender	0.0151
TOLD	LE_FL * DisabilityProxyFall	0.0029
TOLD	Curiosity Corner_FL * MaternalEducation	0.0012
TOLD	Curiosity Corner_NJ * MaternalEducation	0.0055

See notes at end of table.

Table B-16. Significant treatment by covariate interactions from check on homogeneity of regression assumption—Continued

Variable	Effect	Prob F
CMA-A Mathematics Composite	Pre-K Math_CA * Child'sGender	0.0158
CMA-A Mathematics Composite	Pre-K Math_CA * DisabilityProxyFall	0.0102
CMA-A Mathematics Composite	Pre-K Math_NY * Child'sGender	0.0114
CMA-A Mathematics Composite	CC(UNC)_NC * Child'sAge	0.0461
CMA-A Mathematics Composite	CC (UNC)_NC * DisabilityProxyFall	0.0153
CMA-A Mathematics Composite	LB_TX * DisabilityProxyFall	0.0079
CMA-A Mathematics Composite	PA_WI * Child'sGender	0.0237
CMA-A Mathematics Composite	LE_FL * Child'sAge	0.0005
CMA-A Mathematics Composite	LE_FL * DisabilityProxyFall	0.0052
CMA-A Mathematics Composite	DLM with OC_FL * Child'sAge	0.0241
CMA-A Mathematics Composite	Curiosity Corner_KS * DisabilityProxyFall	0.0052
CMA-A Mathematics Composite	Curiosity Corner_NJ * MaternalEducation	0.0414

[1] Building Blocks, Shape Composition task

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

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

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

Example for a single intervention at a single site

An example of testing intervention differences using the linear spline model for child outcomes is described here for the University of New Hampshire research team that implemented one curriculum (*Creative Curriculum* with *Ladders to Literacy*) at one site (NH). In this example, y is the child outcome under modeling; Time_1 is the time lapses between the start of treatment and any subsequent assessment used in the model; Time_2 is the time lapse between the pre-kindergarten spring assessment and later assessments, respectively, as described in the section Repeated Measures Models; and $\text{CRTL} - \text{NH}$ is the intervention by site indicator.

$$\begin{aligned} E(y) = {} & \beta_0 + \beta_1 \text{Time}_1 + \beta_2 \text{Time}_2 + \beta_3 (\text{CTRL-NH}) \\ & + \beta_4 (\text{CRTL-NH} \times \text{Time}_1) + \beta_5 (\text{CTRL-NH} \times \text{Time}_2) \\ & + \beta_6 \text{Gender} + \beta_7 \text{Age} + \beta_8 \text{Race} + \beta_9 \text{Disability} \\ & + \beta_{10} \text{MaternalEducation} \end{aligned}$$

The evaluation includes estimating means for the intervention and control classrooms at a particular time, typically spring of pre-kindergarten year. The intervention by site indicators are coded in such a way that linear contrasts can be conveniently constructed to allow comparisons of groups means and slopes at various time points. With this coding scheme, the group means, adjusted for covariates, are calculated using site-specific intervention group codes of +.5 for intervention and -.5 for control as follows:

$$E(y_{\text{intervention}}) = \beta_0 + \beta_1 \text{Time}_1 + \beta_2 \text{Time}_2 + \beta_3(0.5) \\ + \beta_4(0.5 * \text{Time}_1) + \beta_5(0.5 * \text{Time}_2) + Xy$$

and

$$E(y_{\text{control}}) = \beta_0 + \beta_1 \text{Time}_1 + \beta_2 \text{Time}_2 + \beta_3(-0.5) \\ + \beta_4(-0.5 * \text{Time}_1) + \beta_5(-0.5 * \text{Time}_2) + Xy$$

where X represents the collection of covariates (gender,…, maternal education) and y is their set of β s. Noting that $\text{Time}_2 = 0$ for spring pre-kindergarten scores, the difference between intervention and control group means then would be

$$E(y_{\text{intervention}}) - E(y_{\text{control}}) = \left\{ \frac{\beta_3 + \beta_4 \text{Time}_1, \text{ for Pre} - \text{K Spring scores}}{\beta_3 + \beta_4 \text{Time}_1 + \beta_5 \text{Time}_2, \text{ for K Spring scores}} \right\}$$

Estimates for the slope for the pre-kindergarten portion of the linear spline model are constructed through linear combinations (*C*) of the fixed parameters:

$$(C'\beta)_{\text{slope,Pre-K,intervention}} = \beta_1 + \beta_4(0.5)$$
$$(C'\beta)_{\text{slope,Pre-K,control}} = \beta_1 + \beta_4(-0.5)$$

The difference between intervention and control group slopes for this portion would be

$$(C'\beta)_{\text{slope,Pre-K,intervention}} - (C'\beta)_{\text{slope,Pre-K,control}} = \beta_4$$

Similarly, estimates for the slope for the kindergarten portion of the linear spline model are

$$(C'\beta)_{\text{slope,K,intervention}} = \beta_1 + \beta_2 + \beta_4(0.5) + \beta_5(0.5)$$
$$(C'\beta)_{\text{slope,K,control}} = \beta_1 + \beta_2 + \beta_4(-0.5) + \beta_5(-0.5)$$

Again, the difference between intervention and control group slopes for the kindergarten portion would be

$$(C'\beta)_{\text{slope,K,intervention}} - (C'\beta)_{\text{slope,K,control}} = \beta_4 + \beta_5$$

Example for a double intervention at a single site

A more complicated example of testing intervention differences using the linear spline model for child outcomes is described here for the University of Texas-Houston research team that implemented two intervention curricula (*Doors to Discovery* and *Lets Begin with the Letter People*) at one site (University of Texas Health-Houston) using one shared control group. The model is as follows:

$$\begin{aligned} E(y) = {} & \beta_0 + \beta_1 \text{Time}_1 + \beta_2 \text{Time}_2 + \beta_3(\text{LETP-TX}) \\ & + \beta_4(\text{LETP-TX} \times \text{Time}_1) + \beta_5(\text{LETP-TX} \times \text{Time}_2) \\ & + \beta_6(\text{DOOR-TX}) + \beta_7(\text{DOOR-TX} \times \text{Time}_1) \\ & + \beta_8(\text{DOOR-TX} \times \text{Time}_2) \\ & + \beta_9 \text{Gender} + \beta_{10} \text{Age} + \beta_{11} \text{Race} + \beta_{12} \text{Disability} \\ & + \beta_{13} \text{MaternalEducation} \end{aligned}$$

Note that there are two interventions and one control group at this site. With the appropriate coding scheme, the *Let's Begin with Letter People* (LETP) intervention and shared control group means, adjusted for covariates, are as follows:

$$E(y_{\text{intervention-LETP-TX}}) = \beta_0 + \beta_1 \text{Time}_1 + \beta_2 \text{Time}_2 + \beta_3 (1.0) \\ + \beta_4 (1.0 \times \text{Time}_1) + \beta_5 (1.0 \times \text{Time}_2) + X\gamma$$

and

$$E(y_{\text{control}}) = \beta_0 + \beta_1 \text{Time}_1 + \beta_2 \text{Time}_2 + \beta_3 (\text{-}1.0) \\ + \beta_4 (\text{-}1.0 \times \text{Time}_1) + \beta_5 (\text{-}1.0 \times \text{Time}_2) \\ + \beta_6 (\text{-}1.0) + \beta_7 (\text{-}1.0 \times \text{Time}_1) + \beta_8 (\text{-}1.0 \times \text{Time}_2) + X\gamma$$

Noting that $\text{Time}_2 = 0$ for spring pre-kindergarten scores, the difference between the LETP intervention and shared control group means then would be:

$$E(y_{\text{intervention}-\text{LETP}-\text{TX}}) - E(y_{\text{control}}) = \begin{cases} \beta_3 (2.0) + \beta_4 (2.0 \times \text{Time}_1) + \\ \beta_6 (1.0) + \beta_7 (1.0 \times \text{Time}_1), \text{for Pre - K Spring scores} \\ \beta_3 (2.0) + \beta_4 (2.0 \times \text{Time}_1) + \beta_5 (2.0 \times \text{Time}_2) + \\ \beta_6 (1.0) + \beta_7 (1.0 \times \text{Time}_1) + \beta_8 (1.0 \times \text{Time}_2), \text{for K Spring scores} \end{cases}$$

Estimates for the slope for the pre-kindergarten portion of the linear spline model are

$$(C'\beta)_{\text{slope,Pre-K,intervention-LETP-TX}} = \beta_1 + \beta_4 (1.0)$$

$$(C'\beta)_{\text{slope,Pre-K,control}} = \beta_1 + \beta_4 (-1.0) + \beta_7 (-1.0)$$

The difference between LETP intervention and control group slopes for this portion would be

$$(C'\beta)_{\text{slope,Pre-K,intervention-LETP-TX}} - (C'\beta)_{\text{slope,Pre-K,control}} = \beta_4 (2.0) + \beta_7$$

Similarly, estimates for the slope for the kindergarten portion of the linear spline model are

$$(C'\beta)_{\text{slope,K,intervention-LETP-TX}} = \beta_1 + \beta_2 + \beta_4 (1.0) + \beta_5 (1.0)$$

$$(C'\beta)_{\text{slope,K,control}} = \beta_1 + \beta_2 + \beta_4 (-1.0) + \beta_5 (-1.0) + \beta_7 (-1.0) + \beta_8 (-1.0)$$

Again, the difference between LETP intervention and control group slopes for the kindergarten portion would be

$$(C'\beta)_{\text{slope,K,intervention-LETP-TX}} - (C'\beta)_{\text{slope,K,control}} = \beta_4 (2.0) + \beta_5 (2.0) + \beta_7 + \beta_8$$

Other Modifications

An example for a research team working at more than one site is a relatively straightforward modification: simply add covariates and covariate by time interactions for each site. For a research team working at two sites using one curriculum (e.g., the University of North Carolina working in both North Carolina and Georgia with *Creative Curriculum*), the following model would apply:

$$
\begin{aligned}
E(y) = {} & \beta_0 + \beta_1 \text{Time}_1 + \beta_2 \text{Time}_2 + \\
& \beta_3(\text{CRT\_NC}) + \beta_4(\text{CRT\_NCxTime}_1) + \beta_5(\text{CRT\_NCxTime}_2) + \\
& \beta_6(\text{CRT\_GA}) + \beta_7(\text{CRT\_GAxTime}_1) + \beta_7(\text{CRT\_GAxTime}_2) + \\
& \beta_9 \text{Gender} + \beta_{10} \text{Age} + \beta_{11} \text{Race} + \beta_{12} \text{Disability} + \beta_{13} \text{MaternalEducation}
\end{aligned}
$$

For a research team using three sites (e.g., the Success for All Foundation) the above model is used with the addition of a line for a third site.

The above model can be used to illustrate the modifications made for the other two types of models (simple repeated measures and ANOVA). For the repeated measures model, drop the Time_2 terms creating the modified model:

$$
\begin{aligned}
E(y) = {} & \beta_0 + \beta_1 \text{Time}_1 \\
& \beta_2(\text{CRT\_NC}) + \beta_3(\text{CRT\_NCxTime}_1) + \\
& \beta_4(\text{CRT\_GA}) + \beta_5(\text{CRT\_GAxTime}_1) + \\
& \beta_6 \text{Gender} + \beta_7 \text{Age} + \beta_8 \text{Race} + \beta_9 \text{Disability} + \beta_{10} \text{MaternalEducation}
\end{aligned}
$$

For the ANCOVA model further drop the Time_1 terms and add the baseline measure of y, creating the modified model:

$$
\begin{aligned}
E(y) = {} & \beta_0 + \beta_1(\text{y}_{\text{baseline}}) \\
& \beta_2(\text{CRT\_NC}) + \\
& \beta_3(\text{CRT\_GA}) + \\
& \beta_4 \text{Gender} + \beta_5 \text{Age} + \beta_6 \text{Race} + \beta_7 \text{Disability} + \beta_8 \text{MaternalEducation}
\end{aligned}
$$

Specific Time Points Used in Tests of Impact

As indicated above, the time points of interest for testing of the intervention and control group differences were start of treatment, fall pre-kindergarten, spring pre-kindergarten, and kindergarten assessments. The time points used in impact testing were mean observed values for these assessments. More specifically, the mean values across all the classrooms (treatment and control classrooms) that were used in the model contrasts, estimating group means, and slope and impacts at various points, for Time_1 were as follows: 0.0 for start of treatment; 6.608 weeks for fall pre-kindergarten assessment; 35.376 weeks for spring pre-kindergarten assessment (Time_1); and 87.454 weeks for spring kindergarten assessment. Corresponding Time_2 values used are 0's (zeros) for the pre-kindergarten year and 52.078 for kindergarten spring estimates. The estimated trajectories whose coefficients were used in the test contrasts were estimated by modeling the outcomes using Time_1 and Time_2 values derived from the *actual* specific classroom assessment (start of assessments time values averaged across all classrooms) and start of intervention dates, which varied from classroom to classroom.

Method of Estimating Degrees of Freedom

The use of normal and chi-squared distributions in testing ML estimates of regression coefficients or contrasts is generally too liberal when sample sizes are small. To use the t or F distributions requires specifying denominator degrees of freedom, which is not easy to determine in unbalanced data. To accommodate this, several approaches have been developed to approximate the denominator degrees of freedom. Options include the containment, residual, Satterthwaite, and Kenward-Roger methods. Researchers

(Kenward and Roger 1997, Keselman et al. 1998, Schaalje, McBride, and Fellingham 2002, and Gomez, Schaalje, and Fellingham 2005) have indicated that the Kenward-Roger adjustment provides the most unbiased Type I error rate for complicated covariance structures and small sample sizes. Disparities between the use of different methodologies is generally small. For these analyses, all tests of fixed effects and contrasts (i.e., intervention and control group means at various time points, estimated impacts, various slopes, and impacts on slopes) were tested using the Kenward-Roger method for estimating denominator degrees of freedom.