

**Longitudinal Effects of the  
Arkansas Better Chance Program:  
Findings from First Grade through  
Fourth Grade**

Kwanghee Jung, Ph.D.

W. Steven Barnett, Ph.D.

The National Institute for Early Education Research  
Rutgers University

Jason T. Hustedt, Ph.D.

University of Delaware

Jessica Francis, M.S.

The National Institute for Early Education Research  
Rutgers University

May 2013

Funding for this project was provided by the State of Arkansas and the Pew Charitable Trusts.

### **Acknowledgements**

Many thanks are due to our colleagues at Partners for Inclusive Communities at the University of Arkansas for Medical Sciences (UAMS), who worked with us to carry out this study: Dr. Leanne Whiteside-Mansell, Yvonne Bradshaw, and their field staff. We greatly appreciate assistance from Tonya Russell, Jamie Morrison, Paige Cox, and Paul Lazenby at the Arkansas Division of Child Care and Early Childhood Education, as well as the many school administrators and teachers in Arkansas who made this research possible. Thanks to Dr. Ellen Frede for her ongoing advice on conducting a longitudinal study of state pre-K, and to Matt Burke, Rachel Sansanelli, and Amanda Colon for data management and coordination at NIEER. Finally, we are especially grateful to the children who took time out of their school days to participate in our study.

## Executive Summary

The National Institute for Early Education Research (NIEER) at Rutgers University conducted a longitudinal study of the Arkansas Better Chance (ABC) prekindergarten program to estimate the effects of state-funded pre-K in Arkansas on children's language, mathematics, and literacy skills through fourth grade. Our study has two complementary components: a conventional statistical comparison of ABC participants with children who did not participate in pre-K, as well as a rigorous regression discontinuity design (RDD) that estimates the effects of ABC participation at kindergarten entry. The RDD is the strongest approach for producing unbiased estimates when a randomized trial is not available. RDD findings presented in an earlier report show that ABC had positive impacts on children's language, math, and early literacy skills at kindergarten entry. These impacts were slightly greater than average impacts found by RDD analyses that were conducted across eight states.

As the RDD approach cannot be used for a longitudinal study, it is also necessary to compare children who attended ABC with children from the same age cohorts who did not attend ABC, on a longitudinal basis. We refer to this approach as the "within-cohort analysis." This is the best approach given the data available, but it may be biased by parental choice about whether or not to enroll a child, which is not a problem that affects the RDD approach. We can assess the actual degree of bias by comparing results at kindergarten entry from the within-cohort analysis with RDD results. At kindergarten entry we found the "true" impact, as measured by the RDD, on language to be about 1.3 to 1.5 times that found by the longitudinal study; we also found the "true" impact on math to be about 1.5 to 1.8 times that found by the longitudinal study. In other words, our within-cohort analysis appears to significantly underestimate the impacts of the ABC program. The true impacts of ABC are likely to be about 50 percent larger than (1.3 to 1.8 times) the estimates obtained using a within-cohort approach. At the end of this report we provide adjusted estimates which assume that the degree of bias found with data from the first year is sustained throughout subsequent years, but there is no truly accurate way to make such adjustments. As the differences that led to the initial bias continue to produce differences in children's learning and development over time, we expect the bias, and, therefore, the amount of underestimation, to increase as well.

A report that detailed findings from the first two years of the longitudinal study was completed in 2008. The 2008 report presented data at three key points in time (the beginning and end of kindergarten, and the end of first grade), which indicated that classroom quality in the ABC program was good on average, producing gains in children's language, mathematics, and literacy scores. This report provides estimates of the impact of ABC as children progressed through the 2009-2010 school year. We present results that answer two questions:

(1) What is the impact of ABC regardless of whether the comparison group attended another program (including Head Start and private preschool)?

(2) What is the impact of ABC compared to not attending any center-based preschool at age 4?

Positive effects were found at the end of first and second grade for language, math, and literacy, and at the end of third grade for literacy. These effects are more pronounced when including only children who did not attend another preschool program in the comparison group than when additionally including children who attended a preschool program other than the ABC initiative. One explanation for estimated effects falling off at the end of third grade is provided by another important finding from this study: children who attended ABC were less likely to be retained in grade. This is a key indication that schools are expending extra effort to help those most behind catch up, which disproportionately helps children who did not attend the ABC program. While effective, these efforts are expensive, and may include extra teacher time in the classroom, remedial programs, and even special education. These efforts may gradually reduce the test score advantages for ABC children in later years, but at a substantial cost. Finally, we caution that our longitudinal within-cohort approach considerably underestimates long-term effects, but that ultimately the limitations of this approach are a problem that is difficult to fix when a randomized trial is not possible.

## Introduction

As state prekindergarten initiatives have expanded to serve more and more of the nation's children during the years before kindergarten (Barnett, Carolan, Fitzgerald, & Squires, 2011), it has become especially important to study the effectiveness of such programs. Arkansas is one of the states where pre-K enrollment has expanded most steadily in recent years. Arkansas first began offering state-funded prekindergarten in 1991 through the Arkansas Better Chance Program (ABC), and the state deepened its commitment with new funding for the Arkansas Better Chance for School Success initiative in 2004. After these additional funds became available, state pre-K enrollment increased from 3,104 children enrolled in center-based programs during the 2003-2004 school year to 13,617 children in the 2006-2007 school year (Barnett, Hustedt, Friedman, Stevenson Boyd, & Ainsworth, 2007; Barnett, Hustedt, Robin, & Schulman, 2005). By the 2010-2011 school year, 44% of the state's 4-year olds were enrolled (Barnett et al., 2011).

The majority of ABC participants are served in public schools, though programs also operate in other types of locations such as educational cooperatives, Head Start facilities, and private child care facilities. The ABC prekindergarten initiative originally required lead teachers in each classroom to have a BA or AA degree and specialized training in pre-K education, and assistant teachers to have a CDA credential. As of 2007, lead teachers in public school settings were required to have a bachelor's degree in early childhood education or child development with P-4 certification. In all other settings, one teacher for every three classrooms must have a bachelor's degree in early childhood education, child development, or equivalent. The ABC initiative also features a maximum class size of 20 children with staff-child ratios of 1:10, and offers a number of comprehensive child and family support services. Total state spending during the 2010-2011 school year was approximately \$111 million for center-based programs, and state spending per child enrolled averaged \$5,021. Additional monetary contributions at the local level are also required, and ABC had a total spending level of \$8,126 per child enrolled when accounting for spending from all sources (Barnett et al., 2011).

The National Institute for Early Education Research (NIEER) at Rutgers University carried out a longitudinal study of the Arkansas Better Chance program, with funding from the State of Arkansas. The goal of this study was to estimate the effects of the ABC program, including the extent to which initial benefits result in persistent educational advantages. A substantial research literature indicates that high-quality preschool education not only produces immediate gains but also confers lasting advantages, most often reduced special education placements and grade repetition but also achievement test score gains. A longstanding concern has been that the quasi-experimental approaches these studies have typically used are subject to a problem called "selection bias" that causes them to underestimate effects. Selection bias is the distortion of program effect estimates that occurs when the effects of the characteristics of the population served and the effects of the program cannot be clearly separated. Most concerning are unmeasured characteristics of families (e.g., living in an area of concentrated poverty,

parental attitudes toward education, parental education levels, children's academic potential) that differ between participants and non-participants and may distort the estimates of program effects. Selection bias is possible whenever eligibility criteria or program administrators determine who participates in a program or whenever the eligible population has any control over whether they participate. These conditions are referred to as administrative selection and self-selection, respectively. When estimating the effects of targeted state preschool programs, both administrative selection and self-selection have the potential to bias the estimated effects.

Where feasible, a randomized trial in which children are assigned by lottery to either a treatment or a control condition is the optimal solution to the problem of selection bias. Such an approach eliminates the threat from even unanticipated and unmeasured differences between children in preschool and control groups. However, when this study began, it was judged that a random assignment study of children eligible for ABC was not feasible. Thus, we took an alternative two-pronged approach to control for selection bias. One component of this approach is a conventional statistical comparison of ABC children and preschool nonparticipants over time. We can anticipate that selection bias would be present in such a comparison, but there is no way to predict in advance whether selection bias will have a small or large effect on the results. In fact, the only way to assess the extent of the bias problem is to implement a second component of the study using another research method that avoids selection bias. Hence, this project employed a second research component from the start: a regression discontinuity design (RDD) that, under reasonable assumptions, avoids selection bias.

The RDD approach seeks to eliminate selection bias by comparing children who have enrolled in ABC at program entry with the previous year's ABC pre-K graduates at kindergarten entry. However, the RDD approach cannot provide an estimate of effects beyond kindergarten entry. Therefore, we further compared children who attended ABC with children in the same age cohorts who did not attend ABC, at annual intervals, into their early elementary school years. We refer to this longitudinal approach as the *within-cohort analysis*. By using both RDD and longitudinal methods to estimate the effects of ABC at kindergarten entry, we can judge the extent to which selection bias is a problem by comparing results from the conventional approach to more accurate RDD estimates. If the initial estimates from both analyses are similar, then we can be assured that selection bias is minor and unlikely to affect the longitudinal results. If there is a discrepancy, then we have a measure of the direction and magnitude of the bias in the longitudinal estimates. With this information we can adjust the statistical analyses to reduce the bias. This may or may not be very effective. At the very least, we can say how far the conventional estimates of long-term effects are likely to fall below the true effects of the program.

In Fall 2005, at the outset of this study, we randomly selected two groups of children as our initial sample: (1) a group of approximately 1,000 children who were beginning kindergarten, evenly divided between children who had participated in the ABC program the previous year and children who had not; and (2) a group of approximately 500 4-year-olds who

were just beginning the ABC program. In Fall 2006, we randomly selected approximately 500 additional entering kindergarteners who had not attended the ABC program and added them to this second group. We refer to the older, first group as “Cohort 1,” and to the younger, second group as “Cohort 2.” Children from Cohort 1 were 5 years old and eligible for kindergarten during the 2005-2006 school year, and were typically in fourth grade at the conclusion of our study during the 2009-2010 school year. Children from Cohort 2 were 4 years old and of prekindergarten age during 2005-2006, and were typically in third grade at the end of the study. For RDD analyses, we used data collected only from the approximately 1,000 children (500 from each cohort) who had participated in the ABC program. For within-cohort longitudinal analyses, data from all children were used, and ABC participants from both cohorts were compared with ABC non-participants from both cohorts. Table 1 shows the longitudinal assessment schedule.

Table 1. Scheduled Assessment Periods for Longitudinal ABC Study

Year of Longitudinal Study	Assessment Schedule
Year 1 (2005-06) <i>Cohort 1: Kindergarten</i> <i>Cohort 2: Pre-K</i>	Fall (all children)  Spring (Cohort 1 only)
Year 2 (2006-07) <i>Cohort 1: Grade 1</i> <i>Cohort 2: Kindergarten</i>	Fall (Cohort 2 only)  Spring (all children)
Year 3 (2007-08) <i>Cohort 1: Grade 2</i> <i>Cohort 2: Grade 1</i>	Spring (all children)
Year 4 (2008-09) <i>Cohort 1: Grade 3</i> <i>Cohort 2: Grade 2</i>	Spring (all children)
Year 5 (2009-10) <i>Cohort 1: Grade 4</i> <i>Cohort 2: Grade 3</i>	Spring (all children)

Our study design allows us to examine the effects of the ABC program at several points in time, and from different methodological perspectives. As detailed in an earlier report (Hustedt, Barnett, Jung, & Thomas, 2007), the initial research question asked whether participating in one year of the state-funded ABC preschool program at age 4 has an impact on children's academic skills when they enter kindergarten. We used the RDD approach to answer that question, comparing children who had just started ABC with those who had just completed the program.

This research design has been used previously to examine the effects of state prekindergarten in Oklahoma (Gormley, Gayer, Phillips, & Dawson, 2005) as well as in a number of other states (Frede, Jung, Barnett, Lamy, & Figueras, 2007; Hustedt, Barnett, & Jung, 2007; Wong, Cook, Barnett, & Jung, 2008).

Results from the RDD analyses show that the ABC program has positive—and statistically significant—impacts on children's early language, literacy, and mathematics development (Hustedt et al., 2007). In those earlier analyses, we estimated that attending the ABC program at age 4 yields 31% more growth in children's vocabulary at kindergarten entry, compared to non-ABC preschool education experiences. Children who participated in ABC scored higher on a test of their early math skills—with 37% more growth at kindergarten entry. The ABC program also had large effects on children's understanding of print concepts, more than doubling growth over the year (116%) in print awareness scores.

In the current report we present new results from the longitudinal, within-cohort analysis component of the study. We focus on effects of the ABC program on children's language, literacy, and mathematics skills as measured from first grade to fourth grade years.

### **First through Fourth Grade Outcomes for ABC Participants**

#### The Sample

As described above, two cohorts of children were followed longitudinally to estimate the effects of the ABC program on children's academic skills through fourth grade. At the outset of the study (the 2005-2006 school year) we randomly selected 125 ABC classrooms from a list of all of the state's ABC classrooms. If a single classroom was selected in a particular county, another classroom from an already selected county was substituted, to improve the efficiency of data collection. Approximately 4 ABC participants were selected from each classroom. These children partly comprised Cohort 2. A total of 250 kindergarten classrooms were also randomly selected from the same school districts, and, again, approximately 4 children were selected from each classroom. These children comprised Cohort 1. Twice as many kindergarten classrooms as prekindergarten classrooms were sampled, to ensure that the selected kindergartners would include a group of children who had not attended ABC. During the following school year, a group of entering kindergartners who had not attended ABC was added to Cohort 2 to provide additional children for the comparison group of ABC non-participants.

Research staff from the University of Arkansas for Medical Sciences visited each sampled program site, selected children into the sample using a procedure to ensure randomness, and conducted child assessments. As shown in Table 1, all children were assessed in fall 2005 to provide baseline data, kindergartners were assessed in spring 2006 to provide end-of-year kindergarten data, entering kindergartners were assessed in fall 2006 to provide baseline kindergarten data, and all children were assessed in each spring from 2007 through 2010.

We also supplemented the child assessment data with demographic and family background data. During the initial year of the study, a liaison at each site gathered information on the children’s preschool status, usually from existing school records but occasionally from parent reports. At several additional times, we gathered demographic information from an Arkansas Department of Education school database, the Statewide Information System (SIS). As some relevant data were not available through SIS, we collected further demographic data via phone surveys.

In our first longitudinal report (Hustedt, Barnett, & Jung, 2008), we focused on results from the 2005-2006 and 2006-2007 school years. In the current longitudinal report, we focus on results from spring 2008, 2009, and 2010, as children progressed through elementary school. Table 2 shows sample sizes for the data used in our analyses at each assessment point. We focused our previous report on comparisons between ABC children and those who did not attend any prekindergarten (“No Pre-K”). Data for children who attended non-ABC prekindergarten programs were therefore not included in our previous analyses, and were not reported. In the current report we utilize two different approaches to detect program effects. In the first approach, we combine children from both cohorts who attended non-ABC prekindergarten programs with children who attended no pre-K program at all, and compared them to ABC program participants from both cohorts. We call this approach “ABC v. other.” The second approach recognizes three distinct groups of children from both cohorts: those who attended ABC prekindergarten programs, those who attended non-ABC prekindergarten programs, and those who did not attend any prekindergarten program. This approach allowed us to compare the effects of attending ABC pre-K to attending no pre-K at all. We call this approach “ABC v. no pre-K.” Table 2 provides selected demographic characteristics for these three groups of children at each assessment point.

Table 2. Size of Analyzed Sample, and Demographics at Each Assessment Point

	N	Female (%)	Ethnic Category (%)			Lunch Status (%)			No data
			White/Asian	Black	Hispanic	Free	Reduced	Paid	
<u>2008</u>									
Total	1628	49	63	31	5	35	7	26	32
ABC	783	47	59	35	5	41	8	22	29
No Pre-K	452	53	69	24	7	33	6	25	37
non-ABC Pre-K	393	49	64	32	4	25	7	36	32
<u>2009</u>									
Total	1597	50	64	30	5	35	7	26	32
ABC	764	48	60	34	5	41	8	22	29
No Pre-K	453	52	70	23	7	32	6	24	38
non-ABC Pre-K	380	51	65	31	4	25	7	36	32
<u>2010</u>									

Total	1555	50	64	31	5	40	9	42	9
ABC	745	49	60	35	5	41	10	41	8
No Pre-K	442	52	69	24	7	40	7	41	11
non-ABC Pre-K	368	50	65	30	4	38	7	46	9

### Child Outcome Measures

Child outcome measures in the ABC study focused on receptive vocabulary, mathematics, and literacy skills. Slightly different batteries of measures were used over the course of the study. The differences between these batteries of measures are described in the sections that follow.

#### *Receptive Vocabulary*

Children's receptive vocabulary was measured using the Peabody Picture Vocabulary Test, 3rd Edition (PPVT-III; Dunn & Dunn, 1997) and, for Spanish-speakers, the Test de Vocabulario en Imagenes Peabody (TVIP; Dunn, Padilla, Lugo, & Dunn, 1986). The PPVT is commonly used as quick test of IQ and can be used as a rough assessment of general cognitive abilities. The PPVT is a direct measure of vocabulary size, and the rank order of item difficulties is highly correlated with the frequency with which words are used in spoken and written language. The test is adaptive, establishing a floor below which the child is assumed to know all the answers and a ceiling above which the child is assumed to know none of the answers. Reliability is good as judged by either split-half reliabilities or test-retest reliabilities. The TVIP is appropriate for measuring growth in Spanish vocabulary for bilingual students and for monolingual Spanish speakers.

In the first year of the study, our design specified that bilingual children would be administered assessments in both English and Spanish. All children in the sample were administered the PPVT, regardless of home language, to get a sense of their receptive vocabulary ability in English. All children who spoke some Spanish were also subsequently administered the TVIP. The testing session was then continued, with further measures administered in a single language—English or Spanish—depending upon what the child's teacher designated as his or her best testing language. When running preliminary analyses, for any cases where a child scored better on the TVIP than on the PPVT but the assessor had continued testing in English (or vice versa), we excluded that case from the analyses. During the fall 2006 child assessment period, there were no children in our sample who scored higher on the TVIP than on the PPVT, indicating that the best testing language was English for all children. As a result, we discontinued Spanish-language testing and report only the scores for English-language assessments after 2006. Our analyses focus on raw scores.

#### *Mathematical Skills*

Children's early mathematical skills were measured with three different subtests of the Woodcock-Johnson Tests of Achievement, 3rd Edition (WJ-III; Woodcock, McGrew, & Mather, 2001): Subtest 10 Applied Problems, Subtest 5 Calculation, and Subtest 6 Math Fluency. Subtests 5, 6, and 10 together comprise the **Broad Math Battery** of the Woodcock-Johnson, and test age-appropriate math skills. Subtests of the Woodcock-Johnson are reported to have good reliability. Raw scores are reported.

### *Early Literacy*

During the 2005-2006 school year and in fall 2006, children's early literacy skills were measured using the Print Awareness subtest of the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP; Lonigan, Wagner, Torgeson, & Rashotte, 2002).

The Pre-CTOPPP was removed from our child assessment battery beginning with the spring 2007 assessments, as it no longer tested age-appropriate skills for the children in our sample. Instead, we used additional subtests from the Woodcock-Johnson (Woodcock et al., 2001) to measure early literacy skills. Subtest 1 Letter-Word Identification and Subtest 13 Word Attack were used to assess early literacy among first graders in 2007.

Starting with the spring 2008 assessment period, we dropped Subtest 13 and added Subtest 2 Reading Fluency, and Subtest 9 Passage Comprehension from the Woodcock-Johnson (Woodcock et al., 2001) to measure literacy skills. Subtests 1, 2, and 9 comprise the **Broad Reading Battery** of the Woodcock-Johnson. For all Woodcock-Johnson measures, raw scores are reported.

### Comparisons with Previous Data from the RDD

In our previous report (Hustedt et al., 2008) we presented results from the first two years of the longitudinal component of the Arkansas study. The longitudinal results from each cohort's kindergarten year and Cohort 1's first grade year were compared to the RDD results in order to judge potential bias in the longitudinal estimates. For the current report, we reanalyzed previous longitudinal data from both cohorts' assessments at kindergarten entry in order to maintain consistency with the current analyses, which examine two alternative longitudinal models (ABC v. Other and ABC v. no pre-K) and also include school district in the model. Table 3 compares effect sizes from the longitudinal, within-cohort approach with effect sizes from the RDD analysis.

As stated earlier, our longitudinal methodology may be affected by selection bias due to factors such as parents' decisions about whether to enroll their child in ABC. The RDD methodology corrects for selection bias by necessarily excluding all children who did not participate in the ABC program; however, the RDD approach cannot be used to estimate the unique effects of pre-K after completing later years of school.

The RDD approach produced estimates of the ABC program's impacts at the start of kindergarten year that were substantial in all three areas of learning: receptive vocabulary, early math skills, and print awareness. Table 3 presents estimates of selection bias. Comparison of within-cohort analysis results to RDD results revealed that even by the beginning of kindergarten, the within-cohort analysis underestimated ABC's impacts on children's learning. Despite the evident bias, the within-cohort analysis produced estimates at kindergarten entry that were statistically significant for all 3 measures using data pooled across the two cohorts (except math, when we included children who attended other programs, such as private preschool or Head Start, in our comparison of ABC participants). Unfortunately, the bias appears to be substantial. As discussed in our previous report (Hustedt et al., 2008), this estimated bias is larger than in two other studies that compared RDD and within-cohort estimates for other preschool programs (Tulsa, Oklahoma and New Jersey's low-income urban districts), likely reflecting the wider variation between groups in the ABC study, compared to the populations served in the other two studies.

Table 3. RDD and Conventional Estimates of Effect Size for ABC at Kindergarten Entry

	Receptive		Math		Print Awareness (% Correct)	
Vocabulary						
Regression-discontinuity	0.28*		0.33*		0.82***	
Pooled Cohorts 1 & 2	0.12** <sup>a</sup>	0.20** <sup>b</sup>	0.05 <sup>a</sup>	0.14* <sup>b</sup>	0.15** <sup>a</sup>	0.28** <sup>b</sup>
Estimated Bias	57% <sup>a</sup>	29% <sup>b</sup>	85% <sup>a</sup>	58% <sup>b</sup>	82% <sup>a</sup>	66% <sup>b</sup>

Note. <sup>a</sup>ABC v. Other, <sup>b</sup>ABC v. no pre-K

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

Effect size estimates from the within-cohort analysis approach for receptive vocabulary were 57% or 29% smaller (ABC v. other and ABC v. no pre-K, respectively) than those found by the RDD approach. For math, estimates from the within-cohort analyses were 85% or 58% smaller than those found by the RDD approach. For print awareness, estimates from the within-cohort analyses were 82% or 66% smaller than those found by the RDD approach. There is no reason to expect this discrepancy to improve substantially over time. This should be kept in mind while interpreting results for first grade and beyond. This bias cannot be completely alleviated when a randomized trial is not possible.

In general, we found that estimates of the impacts of pre-K at kindergarten entry, as revealed through RDD findings, are 1.3 to 1.8 times greater than our longitudinal estimates. We now turn to present findings from later years of the study, as children moved from 1<sup>st</sup> grade

through 4<sup>th</sup> grade, again cautioning the reader to interpret the true effects over time as being potentially 1.3 to 1.8 times greater, in light of the limitations of our within-cohort analysis approach.

### **Longitudinal Findings: First through Fourth Grade Child Outcomes**

In our current presentation of the findings of our within-cohort analyses, we focus on child outcome data gathered at four points in children's school careers: the end of the first, second, and third grades for both cohorts, and at the end of Cohort 1's fourth grade year. For each grade, we present data from both cohorts on receptive vocabulary, mathematical skills, and early literacy. Although findings for first grade were presented in a previous report, they have been reanalyzed with models consistent with other grade years and are reviewed again in the current report.

All analyses for the within-cohort approach, as well as our findings, are presented in two ways. The first compares children from both cohorts who attended ABC with a combined group of children who attended preschool programs other than ABC and those who attended no program at all (ABC v. other). The second divides children from both cohorts into three groups, (1) those who attended ABC, (2) those who attended a preschool program other than ABC, and (3) those who attended no program at all (ABC v. no pre-K). For both approaches, we include a number of covariates such as age, gender, ethnicity, and lunch status. We also include school district to account for any unobserved school-level variances directly associated with school districts that might affect children's academic progress.

#### Effect Size for Receptive Vocabulary, Broad Reading, and Broad Math

Next, we present effect sizes of participation in the ABC pre-K program on children's outcome measures. The longitudinal nature of the Arkansas Better Chance evaluation has allowed us to incorporate methodological refinements by strategically combining different subsets of our sample at different points in time. For example, data from Cohort 2 second graders in 2009 can be analyzed together with Cohort 1 second grade data from 2008. This pooled analysis of second grade results is more likely to have enough statistical power to detect the effects of ABC pre-K program participation.

As indicated in table 4, the within-cohort analysis results show that the benefits of participating in the ABC program remain evident through first and second grades, as demonstrated by children's receptive vocabulary, Broad Reading, and Broad Math skills. The estimated effects of ABC on children's receptive vocabulary ranged from 0.05 to 0.14 with an average of 0.10. The estimated effects of ABC pre-K on Broad Reading ranged from 0.06 to 0.15 with an average of 0.10. The estimated effects of ABC on Broad Math ranged from 0 to 0.16 with an average of 0.08. Effects in math were more varied and tended to be smaller.

We found ABC had greater effects when we compared ABC children with children who did not attend another preschool program (ABC v. no pre-K) than when also including children who attended another program (ABC v. other). When interpreting results, it should be noted that the ABC v. other analysis model does not estimate the effects of the ABC pre-K program relative to no-program. Instead, it estimates the effects of these programs relative to other available early care and education alternatives.

The overall pattern of results provides evidence that the ABC pre-K program positively impacts children's outcomes. Statistically significant impacts were found most frequently in earlier grades, and some impacts were sustained several years beyond preschool, despite evidence that the within-cohort analyses underestimated the impacts of ABC.

Table 4. Effect Size

	Grade 1	Grade 2	Grade 3	Grade 4
<b>Receptive Vocabulary</b>				
ABC v. other	0.09 <sup>+</sup>	0.10*	0.05	0.11
ABC v. no pre-K	0.14**	0.14**	0.09	0.14
<b>Broad Reading</b>				
ABC v. other	0.09 <sup>+</sup>	0.08*	0.06	0.09
ABC v. no pre-K	0.15*	0.11**	0.13*	0.12
<b>Broad Math</b>				
ABC v. other	0.02	0.10 <sup>+</sup>	0.00	0.01
ABC v. no pre-K	0.12*	0.16**	0.08	0.03

<sup>+</sup> $p < .10$ , \* $p < .05$ , \*\* $p < .01$

#### Receptive Vocabulary: PPVT

Next, we examine the estimated impacts of ABC pre-K for each content area in which children were assessed. Our first set of analyses examines children's receptive vocabulary at the end of Grades 1, 2, and 3 for both cohorts, and at the end of Grade 4 for Cohort 1. Results from analyses using the two alternative approaches (ABC v. other and ABC v. no pre-K) are summarized in Table 5 and are described in the following sections.

Table 5. ABC Program Effects on Receptive Vocabulary

	Grade 1	Grade 2	Grade 3	Grade 4
--	---------	---------	---------	---------

ABC v. other	1.42 <sup>+</sup>	1.73*	0.93	2.37
ABC v. no pre-K	2.32**	2.61**	1.85	2.83

<sup>+</sup> $p < .10$ , \* $p < .05$ , \*\* $p < .01$ . Program impacts are estimated as increases in PPVT raw scores.

We found that, at the end of the first grade, children in both cohorts who attended the ABC program at age 4 had PPVT scores that were 1.42 ( $p < .10$ , ABC v. other) and 2.32 ( $p < .01$ , ABC v. no pre-K) raw score points higher than children who did not attend pre-K. These represent increases of about 9% (ABC v. other) and 14% (ABC v. no pre-K) of the standard deviation for the control group.

Findings for second grade children in both cohorts also revealed a statistically significant effect of state-funded preschool on receptive vocabulary. Attending an ABC preschool program is associated with an increase of 1.73 ( $p < .05$ , ABC v. other) and 2.61 ( $p < .01$ , ABC v. no pre-K) raw score points on the PPVT at the end of the second grade. These represent improvements of about 10% (ABC v. other) and 14% (ABC v. no pre-K) of the standard deviation for the control group.

At the end of the third grade year, children in both cohorts who attended the ABC preschool program at age 4 had PPVT scores that were 0.93 and 1.85 raw score points higher than children who did not attend ABC pre-K.

We found no significant effect of ABC pre-K on receptive vocabulary scores at the end of the fourth grade year for children in Cohort 1 (children in Cohort 2 were only assessed through their third grade year). Children who participated in ABC at age 4 scored 2.37 (ABC v. other) and 2.83 (ABC v. no pre-K) raw score points higher on the PPVT than children who did not participate in ABC pre-K.

#### Literacy: Broad Reading Battery: Woodcock Johnson Subtests 1, 2, and 9

Our next set of analyses examines children's literacy skills using the Broad Reading Battery of the WJ-III, which is comprised of three subtests: WJ1 Letter-Word Identification, WJ2 Reading Fluency, and WJ9 Passage Comprehension. However, for first grade children in Spring 2007 (Cohort 2), we administered WJ 1 Letter-Word identification and WJ 13 Word Attack instead. These two subtests comprise the Basic Reading Skills battery of the WJ-III; therefore, in this case, we used the Basic Reading Skills battery as the literacy outcome measure (as opposed to the Broad Reading Battery described above).

Results from analyses using the two alternative approaches (ABC v. other and ABC v. no pre-K) are summarized in Table 6 and are described in the following sections.

Table 6. ABC Program Effects on Literacy

	Grade 1	Grade 2	Grade 3	Grade 4
ABC v. other	0.73 <sup>+b</sup>	3.71*	1.26	1.93
ABC v. no pre-K	1.28* <sup>b</sup>	5.17**	2.83*	2.60

<sup>+</sup> $p < .10$ , \* $p < .05$ , \*\* $p < .01$ . Program impacts are estimated as increases in WJ raw scores.

Note: <sup>a</sup> In grade 1 analysis, literacy skills were assessed with WJ1 Letter-Word Identification.

The Woodcock Johnson Letter-Word Identification subtest was used as the literacy measure in our analysis of first grade data for both cohorts. Findings revealed that the estimated effect of state-funded preschool on grade 1 children's literacy skills is 0.73 ( $p < .10$ ) and 1.28 ( $p < .05$ ), for ABC v. other and ABC v. no pre-K respectively. In other words, children who attended the ABC pre-k program at age 4 scored 0.73 (ABC v. other) and 1.28 (ABC v. no pre-K) raw score points higher on the Letter-Word Identification subtest compared to children who did not attend ABC pre-K. The improvement represents about 9% (ABC v. other) and 15% (ABC v. no pre-K) of the standard deviation for the control group.

We collected WJ1 Letter-Word Identification, WJ2 Reading Fluency, and WJ9 Passage Comprehension data for second graders in both cohorts. Therefore, we were able to use the full Broad Reading Battery in all analyses. Findings revealed that the estimated effect of state-funded preschool on grade 2 children's literacy skills is statistically significant. Attending the ABC preschool program at age 4 is associated with increases of 3.71 ( $p < .05$ , ABC v. other) and 5.17 ( $p < .01$ , ABC v. no pre-K) raw score points in Broad Reading. These represent improvements of about 8% (ABC v. other) and 11% (ABC v. no pre-K) of the standard deviation of the control group.

ABC program participation at age 4 was associated with increases of 1.26 ( $p < .05$ , ABC v. other) and 2.83 ( $p < .05$ , ABC v. no pre-K) raw score points on the Broad Reading battery at the end of third grade for both cohorts. The improvements are about 6% (ABC v. other) and 13% (ABC v. no pre-K) of the standard deviation for the control group.

For Cohort 1 children who were in fourth grade in the 2009-2010 school year, the effect of state-funded preschool on children's literacy skills is not statistically significant. When measured at the end of fourth grade, attending the ABC preschool program was associated with increases of 1.93 (ABC v. other) and 2.60 (ABC v. no pre-K) raw scores points on the Broad Reading battery.

#### Mathematics Skills: Broad Math Battery (WJ 5, 6, and 10)

A final set of results from our within-cohort analysis examines children's mathematics skills at the end of grades 1, 2, and 3 for both cohorts, and at the end of grade 4 for Cohort 1.

Findings from two alternative within-cohort analyses (ABC v. other and ABC v. no pre-K) are described below and presented in Table 7.

Table 7. Effects on Mathematic Skills

	Grade 1	Grade 2	Grade 3	Grade 4
ABC v. other	0.28	2.00 <sup>+</sup>	-0.04	0.23
ABC v. no pre-K	2.11*	3.22**	1.88	0.67

<sup>+</sup> $p < .10$ , \* $p < .05$ , \*\* $p < .01$ . Program impacts are estimated as increases in WJ raw scores.

We found that the estimated effect of state-funded preschool on both cohorts' first grade children's math skills is significant when comparing ABC participants to children who did not attend pre-K. Attending the ABC preschool program at age 4 is associated with increases of 0.28 (ABC v. other) and 2.11 ( $p < .05$ , ABC v. no pre-K) raw score points on the Broad Math battery, representing improvements of 2% (ABC v. other) and 12% (ABC v. no pre-K) of the standard deviation for the control group.

At the end of the second grade, children from both cohorts who attended the ABC program at age 4 scored 2.0 ( $p < .10$ , ABC v. other) and 3.22 ( $p < .01$ , ABC v. no pre-K) raw score points higher on the Broad Math battery compared to children who did not attend ABC pre-K. This represents an improvement of about 10% (ABC v. other) and 16% (ABC v. no pre-K) of the standard deviation for the control group.

The effect of the ABC preschool program on children's math skills is not statistically significant at the end of third grade for either cohort. Analysis revealed that the effect of state-funded preschool on children's math skills is related to a decrease of 0.04 (ABC v. other) and an increase of 1.88 (ABC v. no pre-K) raw score points on the Broad Math Battery.

At the end of fourth grade the estimated effect of the ABC preschool Program on Cohort 1 children's mathematical skills is not statistically significant. Children who attended the ABC program at age 4 scored 0.23 (ABC v. other) and 0.67 (ABC v. no pre-K) raw score points higher on the Broad Math battery compared to children who did not attend ABC pre-K. As mentioned previously, fourth grade data are only available for the older children who had reached fourth grade by the conclusion of this study.

#### Adjusting for Selection Bias: Re-examining Results from First Grade through Third Grade

Given that the ABC study includes an RDD component as well as a longitudinal within-cohort component, it is possible to adjust results from the within-cohort approach to more accurately reflect true effects based on findings from the RDD approach. The longitudinal estimates are likely to underestimate the true effects of the ABC program, since, unlike the RDD results, they are subject to selection bias. By comparing results from both approaches at

kindergarten entry – the child assessment point common to both the RDD and within-cohort designs – we can measure the magnitude of selection bias present in our longitudinal results. Overall, we found the within-cohort estimates to be an average of about 63% lower than the RDD estimates from the beginning of kindergarten for both cohorts (see Table 3).

In Table 8, we present current estimates of effect sizes, as well as effect sizes due to selection bias. These results have been mathematically adjusted for the magnitude of selection bias we found to be present at the beginning of kindergarten for both cohorts. Therefore, they likely provide a more accurate representation of the differences in language, literacy, and early math skills between ABC children and children who did not attend a prekindergarten program.

Table 8. Effect Sizes at First Grade, Second Grade, and Third Grade, Adjusted for Selection Bias

	Grade 1	Grade 2	Grade 3
Receptive Vocabulary			
Effect Size (Unadjusted)	0.14**	0.14**	0.09
Effect Size (Adjusted for Selection Bias)	0.23**	0.23**	0.18
Literacy			
Effect Size (Unadjusted)	0.15*	0.11**	0.13*
Effect Size (Adjusted for Selection Bias)	0.24*	0.20**	0.22*
Math			
Effect Size (Unadjusted)	0.12*	0.16**	0.08
Effect Size (Adjusted for Selection Bias)	0.28*	0.32**	0.24

<sup>+</sup> $p < .10$ , \* $p < .05$ , \*\* $p < .01$ . Program impacts are estimated as increases in PPVT raw scores for receptive vocabulary and WJ raw scores for literacy and math.

### Grade Retention

Lastly, we analyzed the effect of the ABC preschool program on children's cumulative grade retention, or the total retention over time that occurred for both cohorts from kindergarten to grade 3. Third grade children who attended the ABC preschool program showed a trend toward lower grade retention ( $p < .10$ ). Overall, cumulative retention rates for children were 10.6% by third grade. The cumulative retention rate was 9.9% for children from both cohorts

who attended ABC pre-K, and 13.3% for those who did not attend any pre-K. Children who attended a non-ABC pre-K program had cumulative retention rates of 8.6%. Overall, the cumulative retention rate among the ABC participants was 3.4 percentage points lower than that of the comparison group who had no preschool program.

Because retention effects are cumulative, they are more likely to be detected as children get older. Grade retention is often associated with a variety of poor academic outcomes including high school dropout. Therefore, the trend toward reduced grade retention by third grade for children who attended high-quality ABC pre-K is an important finding. This finding is particularly noteworthy as it may have an influence on ABC program effects falling off at the end of third grade. That ABC participants were less likely to be retained is a key indication that the schools are making extra efforts to help those most behind catch up, disproportionately helping children who did not attend ABC (and who are more likely to start kindergarten at a disadvantage). While these efforts work, it is an expensive process requiring more teacher time in the classroom, repetition of grades, and special education that the ABC program is designed to reduce. In effect, this catch-up may reduce test score advantages for ABC in later years. In light of the differences in grade retention, coupled with underestimated effects due to selection bias, true long term effects are likely to be considerably larger than the estimates we present in the current report.

### **Summary and Discussion**

The results presented in this report provide evidence of continuing positive impacts of the Arkansas Better Chance Program on children's outcomes. An earlier analysis (Hustedt et al., 2007) using a regression-discontinuity research design found that ABC participants achieved higher vocabulary, mathematics, and print awareness scores when entering kindergarten, compared to children who did not attend ABC pre-K. Our first longitudinal report extended those findings by focusing on a larger sample of children as they progressed through kindergarten and first grade, and also provided information about classroom quality. The current report extends findings through the end of third grade for Cohort 2 and the end of fourth grade for Cohort 1. The long-term effect estimates for ABC v. no pre-K are modest, with effect sizes in the 0.10-0.15 range (0.12-0.25 adjusted for bias) through grade three for both cohorts. Though modest, they are meaningful.

In sum, despite evidence that the longitudinal study underestimates the impacts of ABC, we found continued significant positive effects of the ABC pre-K program on children's receptive vocabulary and math through grade 2 and on literacy through grade 3, for both cohorts. Our ability to find effects through the end of grade 4 was limited because only half of the sampled children (Cohort 1) were old enough to have reached grade 4 by the end of the study. The effects of ABC are stronger when ABC children are compared to children who did not attend another early childhood program, suggesting that Head Start and other preschools also produce achievement gains, though not on average as large as those from ABC.

There are multiple explanations for the apparent decline in effects of ABC as children progress through successive grades. One is that schools are helping children who are behind to catch up, as evidenced by the finding that ABC participants were less likely to be retained at the end of third grade. In addition, we found from the start that our approach underestimated the effects of ABC and this may have worsened over time. Recall that our best estimates of effects were about 50 percent greater at kindergarten entry than estimates using the longitudinal study method.

As was indicated in our 2008 report, the positive effects of ABC pre-K on early literacy skills were non-significant when children were assessed at the end of kindergarten. We suspected such findings were associated with ceiling effects on the Pre-CTOPPP (Lonigan, Wagner, Torgeson, & Rashotte, 2002), which measures basic early reading skills, such as letter and sound recognition and basic print concepts. It was found that most children, regardless of ABC attendance or not, had mastered the tasks presented in the Pre-CTOPPP, and it was necessary to adopt another test to appropriately measure children's progress. As a result, we changed literacy measures in 2007 (WJ1 and WJ13) and again in 2008 (WJ1, WJ2, and WJ 9), to provide more age-appropriate measures, to better detect differences between groups. Since we were unable to detect any significant effects of ABC pre-K on literacy skills at the end of kindergarten due to the limitations of the measure, the positive effects found at the end of grades 1, 2, and 3 do not indicate any change in program impact, but rather the use of a more appropriate measure.

The finding that children who attended ABC pre-K were less likely to have been retained by the end of third grade than those who did not attend any pre-K deserves attention. The end of third grade is early to detect an effect in cumulative measures of school success. As children move forward, it is also possible that other effects will be found on special education, disciplinary problems, and other aspects of school success. The slight differences from other programs suggest that ABC and other programs have similar impacts on school progress. However, it must be acknowledged that unmeasured differences between ABC children and those attending other programs might be masking differences in these outcomes.

As mentioned previously, the effects of ABC participation might be underestimated due to differences among families that led to the choice of ABC, some other program, or no program at all. The limited information we have on family characteristics restricts our ability to carefully model these choices and control for the influences of these other family characteristics on child outcomes. Since the Arkansas Better Chance and Arkansas Better Chance for School Success programs target children from low-income families, this raises the possibility that the many children who attended ABC and are more disadvantaged than those not attending pre-K or attending private pre-K. Specifically, children in the ABC group may have been more disadvantaged overall by virtue of meeting the eligibility requirements of the state prekindergarten program. To address this possibility, family demographic data were gathered from the Arkansas Department of Education's Student Information System in early 2007. We controlled for children's free-and reduced-price lunch status in both the 2008 report and the

current analysis, but data on children's lunch status were still not available for some children and this keeps us from dealing with the problem as well as we might.

It is important to recognize that the Arkansas Better Chance program has vastly expanded since this study began in 2005, and it is now much more accessible to children and families across the State of Arkansas. In order to gather data that are more relevant to the current scope of the ABC program, we have collected new data and continue to analyze the impacts of ABC pre-K at kindergarten entry during the 2010-2011 and 2011-2012 school years. Conducting this new study that draws upon the current population of ABC preschool participants will allow us to gain a better understanding of the impact of the ABC program as it exists now. In addition, the new study will have an enhanced collection of demographic data that should make it possible to produce even better estimates of the ABC program's impacts.

## References

- Barnett, W.S., Carolan, M., Fitzgerald, J., Squires, J. H. (2011). *The state of preschool 2011: State preschool yearbook*. New Brunswick, NJ: National Institute for Early Education Research, Rutgers University.
- Barnett, W. S., Hustedt, J. T., Friedman, A. H., Stevenson Boyd, J., & Ainsworth, P. (2007). *The state of preschool 2007: State preschool yearbook*. New Brunswick, NJ: National Institute for Early Education Research, Rutgers University.
- Barnett, W. S., Hustedt, J. T., Robin, K. B., & Schulman, K. L. (2005). *The state of preschool: 2005 state preschool yearbook*. New Brunswick, NJ: National Institute for Early Education Research, Rutgers University.
- Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83, 596-610.
- Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test-Third Edition (PPVTIII)*. Circle Pines, MN: AGS Publishing.
- Dunn L. M., Padilla, E. R., Lugo, D. E., & Dunn, L. M. (1986). *Test de Vocabulario en Imágenes Peabody (TVIP)*. Circle Pines, MN: AGS Publishing.
- Frede, E., Jung, K., Barnett, W. S., Lamy, C. E., & Figueras, A. (2007). *The Abbott Preschool Program Longitudinal Effects Study (APPLES)*. New Brunswick, NJ: National Institute for Early Education Research, Rutgers University.
- Gormley, W.T., Jr., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-K on cognitive development. *Developmental Psychology*, 41, 872-884.
- Hustedt, J.T., Barnett, W.S., and Jung, K (2008). *Longitudinal effects of the Arkansas Better Chance program*. New Brunswick, NJ: National Institute for Early Education Research, Rutgers University.
- Hustedt, J. T., Barnett, W. S., & Jung, K. (2007). *The effects of the New Mexico PreK initiative on young children's school readiness*. New Brunswick, NJ: National Institute for Early Education Research, Rutgers University.
- Hustedt, J. T., Barnett, W. S., Jung, K., & Thomas, J. (2007). *The effects of the Arkansas Better Chance Program on young children's school readiness*. New Brunswick, NJ: National Institute for Early Education Research, Rutgers University.
- Lonigan, C., Wagner, R., Torgeson, J., & Rashotte, C. (2002). Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP). *Department of Psychology*, Florida State University.

NAEYC & NCTM. (2002). *Early childhood mathematics: Promoting good beginnings*. A joint position statement of the National Association for the Education of Young Children

(NAEYC) and the National Council for Teachers of Mathematics (NCTM). Available at: <http://www.naeyc.org/about.positions/psmath.asp> or <http://www.nctm.org/about/content.aspx?id=6352>.

Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state prekindergarten programs. *Journal of Policy Analysis and Management*, 27(1), 122-154.

Woodcock, R. W., & Munoz, A. F. (1990). *Bateria Woodcock-Munoz Pruebas de Aprovechamiento – Revisados*. Itasca, IL: Riverside Publishing.

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

## Appendix

Table A1. PPVT

	Pre-K		No Pre-K		Diff	Effect Size
	N	Mean	N	Mean	Mean	
K entry	716	69.40	423	65.83	3.57	0.27
K end	618	79.57	338	78.32	1.26	0.10
Grade 1	700	92.24	390	90.28	1.96	0.15
Grade 2	710	103.56	406	100.87	2.69	0.17
Grade 3	757	113.65	437	110.96	2.70	0.15
Grade 4 <sup>a</sup>	409	124.86	184	120.83	4.03	0.21

Note. Mean scores adjusted for age, gender, lunch status, and ethnicity by grade. <sup>a</sup>Grade 4 mean scores are from cohort 1

Table A2. WJ 10: Applied Problems

	Pre-K		No Pre-K		Diff	Effect Size
	N	Mean	N	Mean	Mean	
K entry	729	15.52	438	14.92	0.60	0.17
K end	617	19.95	340	19.81	0.14	0.04
Grade 1	697	25.12	389	24.60	0.52	0.14
Grade 2	708	29.18	402	28.64	0.54	0.14
Grade 3	751	32.70	435	32.43	0.28	0.06
Grade 4 <sup>a</sup>	406	35.79	183	35.63	0.16	0.03

Note. Mean scores adjusted for age, gender, lunch status, and ethnicity by grade. <sup>a</sup>Grade 4 mean scores are from cohort 1

Table A3. WJ1: Letter-Word Identification

	Pre-K		No Pre-K		Diff	Effect Size
	N	Mean	N	Mean	Mean	
K end <sup>a</sup>	264	22.42	198	22.24	0.18	0.03
Grade 1	696	34.32	390	33.11	1.21	0.16
Grade 2	706	42.37	405	41.35	1.02	0.13
Grade 3	753	48.17	429	47.39	0.78	0.09
Grade 4 <sup>b</sup>	407	52.75	184	52.15	0.60	0.06

Note. Mean scores adjusted for age, gender, lunch status, and ethnicity by grade. <sup>a</sup>K end mean scores are from cohort 2. <sup>b</sup>Grade 4 mean scores are from cohort 1

Table A4. WJ2: Reading Fluency

	Pre-K		No Pre-K		Diff	Effect Size
	N	Mean	N	Mean	Mean	
Grade 1 <sup>a</sup>	291	16.80	220	16.10	0.70	0.07
Grade 2	709	29.26	403	28.10	1.16	0.12
Grade 3	757	36.22	438	35.13	1.09	0.11
Grade 4 <sup>b</sup>	406	42.90	184	41.42	1.48	0.13

Note. Mean scores adjusted for age, gender, lunch status, and ethnicity by grade. <sup>a</sup>Grade 1 mean scores are from cohort 2. <sup>b</sup>Grade 4 mean scores are from cohort 1

Table A5. WJ5: Calculation

	Pre-K		No Pre-K		Diff	Effect Size
	N	Mean	N	Mean	Mean	
Grade 1	683	8.81	374	8.38	0.43	0.18
Grade 2	709	11.26	405	11.19	0.07	0.03
Grade 3	757	14.74	437	14.35	0.39	0.12
Grade 4 <sup>a</sup>	400	17.22	179	16.83	0.38	0.10

Note. Mean scores adjusted for age, gender, lunch status, and ethnicity by grade. <sup>a</sup>Grade 4 mean scores are from cohort 1

Table A6. WJ6: Math Fluency

	Pre-K		No Pre-K		Diff	Effect Size
	N	Mean	N	Mean	Mean	
Grade 1	683	26.26	375	24.27	1.98	0.18
Grade 2	709	38.72	407	37.61	1.11	0.08
Grade 3	760	46.84	439	45.98	0.86	0.05
Grade 4 <sup>a</sup>	408	57.80	184	57.35	0.46	0.03

Note. Mean scores adjusted for age, gender, lunch status, and ethnicity by grade. <sup>a</sup>Grade 4 mean scores are from cohort 1

Table A7. WJ9: Passage Comprehension

	Pre-K		No Pre-K		Diff	Effect Size
	N	Mean	N	Mean	Mean	
Grade 1	289	17.22	219	17.20	0.02	0.00
Grade 2	706	23.13	402	22.63	0.51	0.10
Grade 3	752	27.09	436	26.44	0.65	0.13
Grade 4 <sup>a</sup>	404	29.89	184	29.10	0.79	0.17

Note. Mean scores adjusted for age, gender, lunch status, and ethnicity by grade. <sup>a</sup>Grade 4 mean scores are from cohort 1

Table A8. Broad Math (WJ 5, WJ6, and WJ10)

	Pre-K		No Pre-K		Diff	Effect Size
	N	Mean	N	Mean	Mean	
Grade 1	701	19.65	392	18.62	1.03	0.19
Grade 2	706	26.40	400	25.81	0.58	0.10
Grade 3	747	31.47	432	31.01	0.46	0.06
Grade 4 <sup>a</sup>	396	36.97	178	36.70	0.28	0.04

Note. Mean scores adjusted for age, gender, lunch status, and ethnicity by grade. <sup>a</sup>Grade 4 mean scores are from cohort 1

Table A9. Broad Reading (WJ 1, WJ2, and WJ9)

	Pre-K		No Pre-K		Diff	Effect Size
	N	Mean	N	Mean	Mean	
Grade 1 <sup>a</sup>	289	22.54	219	22.08	0.46	0.07
Grade 2	700	31.63	397	30.81	0.82	0.12
Grade 3	740	37.19	426	36.40	0.79	0.11
Grade 4 <sup>b</sup>	399	42.00	184	40.86	1.14	0.16

Note. Mean scores adjusted for age, gender, lunch status, and ethnicity by grade. <sup>a</sup>Grade 1 mean scores are from cohort 2. <sup>b</sup>Grade 4 mean scores are from cohort 1