Longitudinal Effects of the Arkansas Better Chance Program: Findings from Kindergarten and First Grade

Jason T. Hustedt, Ph.D.
W. Steven Barnett, Ph.D.
Kwanghee Jung, Ph.D.

The National Institute for Early Education Research
Rutgers University

May 2008

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 partnered with us to carry out this study: Dr. Leanne Whiteside-Mansell, Yvonne Bradshaw, and their field staff. We greatly appreciate assistance from Paul Lazenby, Tonya Russell, and Jamie Morrison 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 at NIEER for her ongoing advice on conducting a longitudinal study of state pre-K, and to Amanda Colon and Jessica Thomas for data management and coordination at NIEER. Finally, we are especially grateful to the kindergartners and first graders who took time out of their school days to participate in our ongoing study.

Executive Summary

During the 2005-2006 school year, the National Institute for Early Education Research (NIEER) at Rutgers University started a 5-year longitudinal study of the Arkansas Better Chance (ABC) prekindergarten program. A goal of this work is 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 children with children who did not participate in pre-K, and a sophisticated regression-discontinuity design (RDD) that estimates the effects of ABC participation at kindergarten entry. The advantage of the RDD is that it eliminates the threat of selection bias, which arises when there are initial differences between children who participate in a preschool program and those who do not. RDD findings presented in an earlier report show that ABC has statistically significant positive impacts on children's language, math, and early literacy skills at kindergarten entry. The RDD approach cannot be used beyond kindergarten entry, however, so it is necessary to also use a conventional approach when examining longer term effects of pre-K. The conventional approach has the advantage of allowing us to estimate the effects of ABC as children progress through elementary school. Data from the RDD and conventional approaches can then be compared to estimate the effects of selection bias on the conventional results.

In the current report, we present data about classroom quality in the ABC program, as classroom quality is likely a key factor through which prekindergarten programs produce positive effects for children. Findings from structured classroom observations using standardized measures of overall quality, classroom support for early literacy, and classroom support for mathematics were generally positive. Overall classroom quality and classroom support for early literacy were good, while scores on classroom mathematics environments suggest that this may be an area to support further.

We also examine children's language, mathematics, and literacy scores at three key points in time: the beginning of kindergarten, the end of kindergarten, and the end of first grade. Our analyses focus primarily on conventional statistical comparisons of ABC participants with children who did not attend pre-K, using data from two cohorts of children. Findings from conventional analyses at the beginning of kindergarten point toward statistically significant positive effects of the ABC program on measures of language, math, and literacy, consistent with the findings from the earlier RDD analyses. However, results from the conventional analyses also showed substantial evidence of selection bias, and underestimate the effects of participating in pre-K.

Statistically significant effects of the ABC program were found for our language measure at the end of the kindergarten year, and for measures of math and early literacy at the end of first grade. These findings are extremely positive considering that the conventional analyses underestimate the effects of ABC participation due to selection bias. Once adjusted for selection bias, estimates of the effects of ABC on children's language, math, and early literacy skills are fairly consistent with the earlier RDD results. Future reports will provide new estimates of the impacts of participating in ABC as children progress through the 2009-2010 school year.

Introduction

As state prekindergarten initiatives have expanded to serve more and more of the nation's children during the years before kindergarten (Barnett, Hustedt, Friedman, Stevenson Boyd, & Ainsworth, 2007), it is 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 et al., 2007; Barnett, Hustedt, Robin, & Schulman, 2005). By 2006-2007, more than 1 in 5 of the state's 4-year-olds was enrolled, and growth continues.

The ABC prekindergarten initiative requires lead teachers in each classroom with a BA or AA degree and specialized training in pre-K education, and assistant teachers with a CDA credential. It also features a maximum class size of 20 with staff-child ratios of 1:10, and offers a number of comprehensive child and family support services. While the majority of ABC participants are served in public schools, programs also operate in other types of locations such as educational cooperatives, Head Start facilities, and private child care facilities. Total state spending during the 2006-2007 school year was approximately $59 million for center-based programs, and state spending per child enrolled averaged $4,316. Additional monetary contributions at the local level are also required, and ABC had a total spending level of $7,194 per child enrolled when spending from all sources was counted (Barnett et al., 2007).

In Fall 2005 the National Institute for Early Education Research (NIEER) at Rutgers University started a 5-year longitudinal study of the Arkansas Better Chance program, with funding from the State of Arkansas. The goal of this study is 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 confers lasting advantages, most often reduced special education placements and grade repetition but also achievement test score gains. A longstanding concern is that the quasi-experimental approaches these studies typically have 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 worrying 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 are referred to as administrative selection and self-selection. 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 confronting the selection bias problem. One component of this approach is a conventional statistical comparison of ABC children and preschool non-participants 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 have 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 this bias by comparing only children who have enrolled in the program, one group at program entry and the other group (last year's pre-K graduates) at kindergarten entry. However, the RDD approach cannot provide an estimate of effects beyond kindergarten entry. A conventional comparison of ABC and non-preschool groups is still necessary to obtain estimates of effects after the beginning of kindergarten.

By using both methods to estimate the effects at kindergarten entry, we can judge the extent of the selection bias problem by comparing results from the conventional approach to the 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 not, then we have a measure of the direction and magnitude of the bias in the longitudinal estimates from the conventional treatment versus the no-treatment comparison group approach. With this information we can try to 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 500 4-year-olds who were just beginning the ABC program, and (2.) a group of 1,000 additional children who were beginning kindergarten. The kindergartners in this initial sample were approximately evenly divided between children who had participated in the ABC program the previous year and children who had not. At the outset of the 2006-2007 school year, we randomly selected an additional 500 entering kindergarteners who had not attended the ABC program.

All participants in the study will be followed through the 2009-2010 school year. Our longitudinal sample includes two cohorts of children: *Cohort 1* includes children who were 5 years old and eligible for kindergarten during the 2005-2006 school year, and *Cohort 2* includes children who were 4 years old and of prekindergarten age during 2005-2006. Children from Cohort 1 will be in fourth grade by the end of the study, while children from Cohort 2 will be in third grade. Table 1 shows the longitudinal assessment schedule for children in Cohorts 1 and 2.

Table 1. Scheduled Assessment Periods for Longitudinal ABC Study

Year of Longitudinal Study	Cohort 1 Assessment Schedule	Cohort 2 Assessment Schedule
Year 1 (2005-06)	Fall of kindergarten year Spring of kindergarten year	Fall of preschool year
Year 2 (2006-07)	Spring of 1st grade year	Fall of kindergarten year (new non-ABC children added) Spring of kindergarten year
Year 3 (2007-08)	Spring of 2nd grade year	Spring of 1st grade year
Year 4 (2008-09)	Spring of 3rd grade year	Spring of 2nd grade year
Year 5 (2009-10)	Spring of 4th grade year	Spring of 3rd grade year

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 was 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 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) and a number of other states (Frede, Jung, Barnett, Lamy, & Figueras, 2007; Hustedt, Barnett, & Jung, 2007; Wong, Cook, Barnett, & Jung, 2008). Further information about RDD approach can be found in the Appendix.

Results from the RDD analyses for Arkansas show that the ABC program has positive—and statistically significant—impacts on children's early language, literacy, and mathematics development (Hustedt et al., 2007). More specifically, attending the ABC program at age 4 yields 31% more growth in children's vocabulary at kindergarten entry, compared to preschool education experiences they would have had without attending ABC. 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 conventional, longitudinal component of the Arkansas study. We focus on effects of the ABC program on children's language, literacy, and mathematics skills as measured during the kindergarten and first grade years. As in a similar NIEER study underway in New Jersey (Frede et al., 2007), the longitudinal results can be compared to the RDD results in order to judge potential bias in the longitudinal estimates. We capitalize on our earlier use of the RDD approach in Arkansas (Hustedt et al., 2007) by interpreting results from the conventional, longitudinal approach in light of the selection bias revealed through comparisons to the RDD findings. Setting the stage for our longitudinal analysis, we begin this report with an overview of classroom quality in ABC

programs, providing details about the types of prekindergarten classroom environments that children in our sample experienced.

Classroom Quality Measures

During the Spring of 2006, we conducted structured observations of ABC classrooms. These observations focused on overall classroom quality, classroom support of reading skills, and classroom support of mathematics skills.

Overall Classroom Quality

Overall program quality was assessed by trained observers using a standardized measure of preschool classroom structure and process, the Early Childhood Environment Rating Scale – Revised (ECERS-R; Harms, Clifford, & Cryer, 2005). This measure has been used extensively in the field and has well-established validity and reliability. The validity of the ECERS-R is supported by high correlations between both the scale items and ratings of items as highly important by a panel of nationally recognized experts, and between scale scores and ratings of classroom quality by experts. Internal consistency as measured by Cronbach's alpha is reported by the authors to be adequate, ranging from .81 to .91. Classroom quality is rated on a 7-point Likert scale, indicating a range of quality from inadequate (1) to excellent (7). The seven ECERS-R subscales are: Space and Furnishings, Personal Care Routines, Language-Reasoning, Activities, Interaction, Program Structure, and Parents and Staff. Average subscale scores are calculated, as well as a total scale score averaged across all 43 items in the scale. The total scale has excellent internal consistency for the current sample ($\alpha = .92$).

Classroom Support of Literacy Skills

The extent to which the classroom environment is supportive of children's literacy development was measured with the Supports for Early Literacy Assessment (SELA; Smith, Davidson, & Weisenfeld, 2001). This measure has been revised for use by NIEER with the deletion of 4 items that overlap with the ECERS-R. The revised measure includes 16 items on a scale from 1 to 5, indicating a range from low quality (1) to high quality (5) for the support of early literacy development. Six subscales are: The Literate Environment, Language Development, Knowledge of Print/Book Concepts, Phonological Awareness, Letters and Words, and Parent Involvement. Internal consistency among scale items as measured by Cronbach's alpha on the current sample is excellent at .92.

Classroom Support of Mathematics Skills

Classroom support for the development of children's early mathematical skills was measured using the Preschool Classroom Mathematics Inventory (PCMI; Frede, Weber, Hornbeck, Stevenson Boyd, & Colon, 2005). This tool measures the materials and strategies used in the classroom to support children's early mathematical concept development, including counting, comparing, estimating, recognizing number symbols, classifying, seriating, geometric shapes, and spatial relations. The standards of the National Council of Teachers of Mathematics and the National Association for the Education of Young Children (2002) inform the PCMI.

This instrument is comprised of 11 items on a 5-point scale—from low quality (1) to high quality (5)—and has two subscales, Materials and Numeracy and Other Mathematical Concepts. Internal consistency among the test items as measured by Cronbach's alpha is good at .86. The PCMI has been found to predict child progress on a standardized math assessment (Frede, Lamy, & Boyd, 2005).

Observer Training Procedures

In order to conduct classroom observations, the University of Arkansas for Medical Sciences hired observers with specific expertise in early childhood education or a related field. Initial training in administering the observation protocols was provided in workshops led by NIEER staff. Trainees then observed in preschool classrooms alongside a trained observer to establish reliability on each observation instrument. Scores of the trainee and the reliable observer were compared, item by item. The true score for each item was determined through discussion but was generally that of the trained observer. A reliability score for the trainee was computed by determining how many exact matches by item she/he had with the true score and how many were only one point above or below the true score. In the NIEER protocol for the ECERS-R, a trainee must complete three observations with 80% or above exact matches or one-away from the true score. The trainee must achieve at least 70% exact agreement for the PCMI and SELA for all three sessions. After five sessions, if the observer is not reliable, he or she is not included in data collection.

Observation Protocol

Classroom observers followed a standard protocol for all observations. Observations should last no less than 3 hours and should include greeting and at least one meal or snack. A brief 15-20 minute teacher interview followed each observation. Teachers were asked questions about typical classroom practices that were not observed, as well as their educational background and teaching experience. When scheduling observations, observers determined if it was likely to be a typical day by asking if field trips, assemblies, or planned absences were scheduled. They did not reveal which teacher would be observed. However, if the teacher had an unplanned absence, the observation was conducted anyway and the interview took place with the assistant teacher. The presence of substitute teachers was noted in the data, but substitute teachers were included in analyses because our goal was to capture children's typical experiences. Having a substitute is one of those experiences. Observers introduced themselves to the classroom staff and briefly explained what they would be doing. They made an effort to be as unobtrusive as possible, and limited conversations with teachers and children to minimize their impact.

Sample for Classroom Observations

In Spring 2006, 68 classrooms were observed in order to gather data on overall classroom quality, classroom support of reading skills, and classroom support of mathematics skills in ABC programs. Almost all of the teachers in these classrooms had either a bachelor's degree (58% of teachers) or a master's degree (36% of teachers), and 86% were certified. ABC teachers in these classrooms had a mean of 4.8 years of experience teaching preschool.

Classroom Quality Findings

Overall Classroom Quality

An important contributing factor to positive outcomes produced by prekindergarten programs is the overall quality of educational services offered in children's classrooms. The seven subscales of the ECERS-R each measure different aspects of overall classroom quality. Table 2 reports average ECERS-R subscale scores and the average total score for the complete sample of ABC classrooms.

The overall mean score on the ECERS-R was 5.26 for the entire sample of classrooms. Of these classrooms, 67.6% had mean scores above 5.0, or in the good to excellent range. Mean scores on the ECERS-R subscales ranged from a low of 4.69 for Activities to a high of 5.92 for Interaction. For most of the subscales there was a wide range of scores across the classrooms in our sample, with some classrooms receiving very high scores and others receiving much lower scores. Four subscales are particularly relevant to the educational effectiveness of the program: Language and Reasoning, Activities, Interaction, and Program Structure. ABC preschool classrooms scored 5 or better on both the Interaction (M = 5.92, SD = 1.21) and Program Structure (M = 5.86, SD = .98) subscales.

Table 2. ECERS-R Total and Subscale Scores

ECERS-R Subscale	Total (N = 68)	
	M (SD)	Range
Space and Furnishings	5.75 (.72)	4.25-7.00
Personal Care Routines	4.94 (1.42)	2.17-7.00
Language-Reasoning	4.90 (1.18)	1.75-7.00
Activities	4.69 (.84)	2.50-6.40
Interaction	5.92 (1.21)	1.80 -7.00
Program Structure	5.86 (.98)	3.33-7.00
Parents and Staff	5.17 (.85)	3.67-7.00
Overall ECERS-R score	5.26 (.70)	3.37-6.52

Classroom Support of Literacy Skills

The SELA measures the classroom environment and teaching practices that lead to early literacy and language development. The average total SELA score across 68 ABC classrooms was 3.28 (SD = .67). On a scale of 1 to 5, with 1 representing very low quality and 5 representing high quality, this score indicates that most ABC preschool classrooms can be characterized as providing good support for children's language and literacy development. Mean

overall SELA scores for ABC classrooms ranged from 1.56 to 4.93. Table 3 reports the SELA average item scores and the average total score for the sample of ABC classrooms.

Table 3. SELA Total and Item Scores

SELA Item	Total (N = 68)	
	M (*SD*)	Range
Item 1. Using print for a purpose	4.21 (.94)	2.0 - 5.0
Item 2. Creating inviting places to look at books	4.41 (.85)	2.0 - 5.0
Item 3. Inviting interest in a wide variety of books	3.75 (1.13)	1.0 - 5.0
Item 4. Writing materials are available and easy to use	4.00 (.85)	2.0 - 5.0
Item 5. Literacy items and props in pretend area	2.87 (1.22)	1.0 - 5.0
Item 6. Encouraging children using oral language	3.44 (1.20)	1.0 - 5.0
Item 7. Introduce new words	2.87 (1.20)	1.0 - 5.0
Item 8. Activities promoting oral language development	3.04 (1.11)	1.0 - 5.0
Item 9. Sharing books	3.96 (.89)	1.0 - 5.0
Item 10. Functions and features of print	2.74 (1.15)	1.0 - 5.0
Item 11. Phonological awareness	2.51 (1.18)	1.0 - 5.0
Item 12. Helping children recognize letters	2.94 (1.22)	1.0 - 5.0
Item 13. Promoting children's interest in writing	3.22 (1.23)	1.0 - 5.0
Item 14. Promoting home based support	2.71 (.92)	1.0 - 5.0
Item 15. Activities to involve parent in children's literacy	2.59 (.90)	1.0 - 5.0
Item 20. Promoting native language	2.27 (1.34)	1.0 - 5.0
Overall SELA Score	3.28 (.67)	1.56 - 4.93

For items on the SELA there was a broad range of scores across different ABC classrooms. On each item, there were classrooms that received the highest possible score of 5. However, on every item there were also classrooms that scored in the lowest quality range, receiving scores of 1 or 2. SELA items with the highest mean scores were "Creating inviting places to look at books" (M = 4.41, SD = .85), "Using print for a purpose" (M = 4.21, SD = .94),

and "Writing materials are available and easy to use" (*M* = 4.00, *SD* = .85). Each of these items scored at least a 4, indicating scores near the ideal. The lowest scoring SELA item was "Promoting native language," with an average score of 2.27 (*SD* = 1.34).

Classroom Support of Mathematics Skills

The PCMI measures the materials and strategies used in classrooms to support children's early mathematical concept development, and is comprised of 11 items on a 5-point scale, from low quality (1) to high quality (5). The mean PCMI total score across the 68 sample classrooms was 2.37 (*SD* = .69). This suggests that the average ABC classroom was providing limited support for the development of children's mathematics skills. Mean overall PCMI scores for ABC classrooms ranged from 1.00 to 4.09. See Table 4 below for the distribution of scores on each of the 11 PCMI items.

Table 4. PCMI Scores in ABC Programs

PCMI Item	Total (*N* = 68)	
	M (*SD*)	Range
Item 1. Materials for counting	3.82 (1.05)	2.0 - 5.0
Item 2. Materials for measuring	3.31 (1.11)	1.0 - 5.0
Item 3. Materials for classifying	3.12 (1.03)	1.0 - 5.0
Item 4. Materials for geometry	3.19 (1.07)	1.0 - 5.0
Item 5. Teachers encourage one-to-one correspondence	1.51 (.99)	1.0 - 5.0
Item 6. Teachers encourage counting	2.22 (1.09)	1.0 - 5.0
Item 7. Teachers encourage estimation	1.90 (1.25)	1.0 - 5.0
Item 8. Teachers use math terminology	1.56 (1.00)	1.0 - 5.0
Item 9. Teachers measure and compare	1.60 (1.00)	1.0 - 5.0
Item 10. Teachers encourage classification	1.94 (1.22)	1.0 - 5.0
Item 11. Teachers encourage geometry	1.82 (1.09)	1.0 - 4.0
Overall PCMI score	2.37 (.69)	1.00 - 4.09

As with the SELA, for items on the PCMI there was a broad range of scores across different ABC classrooms. On 10 of the 11 PCMI items, there were classrooms that received the highest possible score of 5. However, on each item there were also classrooms that scored in the lowest quality range, receiving scores of 1 or 2. The items with the highest mean scores on the

PCMI were "Materials for counting" (*M* = 3.82, *SD* = 1.05), "Materials for measuring" (*M* = 3.31, *SD* = 1.11), "Materials for geometry" (*M* = 3.19, *SD* = 1.07), and "Materials for classifying" (*M* = 3.12, *SD* = 1.03). These items scored 3 or more, indicating fair or adequate quality. The lowest scoring PCMI item was "Teachers encourage one-to-one correspondence," with an average score of 1.51 (*SD* = .99).

Kindergarten and First Grade Outcomes for ABC Participants

The Sample

As described above, our sample includes two cohorts of children who are being followed longitudinally, to estimate the effects of the ABC program on children's academic skills through fourth grade. Our sampling strategy at the outset of the study (the 2005-2006 school year) involved randomly selecting 125 ABC classrooms from a list of all 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. A total of 250 kindergarten classrooms also were randomly selected from the same school districts. Again, approximately 4 children were selected from each classroom. 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. These kindergartners serve as one of the cohorts of children in our longitudinal study. During the 2006-2007 school year, a group of children who had not attended ABC was added to the group of children who attended ABC the previous school year, completing the formation of a second study cohort with both ABC and non-ABC children. In this report, we refer to the children who were of kindergarten age in 2005-2006 as Cohort 1, and refer to the children who were of prekindergarten age that year as Cohort 2.

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. Children in Cohort 1 were assessed in Fall 2005 to provide baseline kindergarten data, in Spring 2006 to provide end-of-year kindergarten data, and again in Spring 2007 to provide data from the end of first grade. Children in Cohort 2 were assessed in Fall 2005 to provide baseline prekindergarten data (only if they were ABC participants), in Fall 2006 to provide baseline kindergarten data, and Spring 2007 to provide end-of-year kindergarten data.

We also collected demographic and family background data. During the 2005-2006 school year, a liaison at each site gathered information on the children's preschool status, usually from existing school records but occasionally from parent report. During the 2006-2007 school year, we gathered additional demographic information from the Arkansas Department of Education's Arkansas Public Schools Computer Network (APSCN).

In this first longitudinal report, we focus on kindergarten and first grade results from the 2005-2006 and 2006-2007 school years. Table 5 shows sample sizes for the data used in our analyses at each assessment point. Because our focus is 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 are not included in our analyses, and are not reported here. Table 5 also provides selected demographic characteristics for children at each assessment point.

Table 5. Size of Analyzed Sample, and Demographics at Each Assessment Point

	N	Female (%)	Ethnic Category (%)				Lunch Status (%)			
			White/Asian	*Black*	*Hispanic*	*Other*	*Free*	*Reduced*	*Paid*	*Missing Data*
C1 K Entry										
Total	**748**	48.3	61.7*	31.1***	6.7	0.4	37.4*	7.9	22.3	32.4**
ABC	**530**	48.9	59.0	35.1	5.5	0.4	40.4	8.7	22.1	28.9
No Pre-K	**218**	46.8	68.3	21.6	9.6	0.5	30.3	6.0	22.9	40.8
C2 K Entry										
Total	**586**	50.2	62.6*	31.2*	4.8	0.5	36.5	5.6	20.3	37.5
ABC	**300**	47.0	58.0	35.3	5.0	0.7	40.3	6.3	17.7	35.7
No Pre-K	**286**	53.5	67.5	26.9	4.5	0.3	32.5	4.9	23.1	39.5
C1 End of K										
Total	**535**	47.7	62.8*	28.2***	8.6*	0.4	38.5	8.8	23.6	29.2
ABC	**395**	48.6	61.3	32.2	6.3	0.3	40.5	9.4	23.3	26.8
No Pre-K	**140**	45.0	67.1	17.1	15.0	0.7	32.9	7.1	24.3	35.7
C2 End of K										
Total	**535**	50.1	64.5**	30.7**	4.7	0.2	36.6	6.0	21.9	35.5
ABC	**301**	47.5	58.5	36.5	4.7	0.3	39.5	6.0	19.6	34.9
No Pre-K	**234**	3.4	72.2	23.1	4.7	0.0	32.9	6.0	24.8	36.3
C1 End of Grade 1										
Total	**641**	49.5	66.0	26.5**	6.9	0.6	41.7	8.9	26.7	22.8*
ABC	**451**	49.9	64.1	29.5	5.5	0.9	43.7	9.8	26.6	20.0
No Pre-K	**190**	48.4	70.5	19.5	0.0	10.0	36.8	6.8	26.8	29.5

Note: C1 represents Cohort 1 and C2 represents Cohort 2. K represents kindergarten.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Child Outcome Measures

Child outcome measures in the ABC study focused on receptive vocabulary, mathematics, and early literacy skills. Slightly different batteries of measures were used during the 2005-2006 and 2006-2007 school years. All 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. Raw scores are reported.

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 have discontinued Spanish-language testing in this study. We report only the scores for English-language assessments starting with the 2006-2007 school year.

Mathematical Skills

Children's early mathematical skills were measured with the Woodcock-Johnson Tests of Achievement, 3rd Edition (WJ-III; Woodcock, McGrew, & Mather, 2001) Subtest 10 Applied Problems. For children whose best testing language was Spanish during the 2005-2006 school year, the *Bateria Woodcock-Munoz Pruebas de Aprovechamiento – Revisado* (Woodcock & Munoz, 1990) *Prueba 25 Problemas Aplicados* was used. Subtests of the Woodcock-Johnson are reported to have good reliability. Raw scores are reported.

Starting with the Spring 2007 assessment period, we added two more Woodcock-Johnson subtests to the assessment battery for first graders: 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.

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 designed as a downward extension of the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgeson, & Rashotte, 1999), which measures phonological sensitivity in elementary school-aged children. Although not yet published when these data were collected, the Pre-CTOPPP has been used with middle-class and low-income samples and includes a Spanish version. Relatively little technical information is available about the performance and psychometric properties of the Pre-CTOPPP.

Print Awareness items from the Pre-CTOPPP measure whether children recognize individual letters and letter-sound correspondences, and whether they differentiate words in print from pictures and other symbols. The percentage of items answered correctly out of the 36 total subtest items is reported.

The Pre-CTOPPP was removed from our child assessment battery beginning with the Spring 2007 assessments, as it was no longer testing age-appropriate skills. Instead, we used additional subtests from the Woodcock-Johnson (Woodcock et al., 2001) to measure early literacy skills. Woodcock-Johnson Subtest 1 Letter-Word Identification and Subtest 21 Sound Awareness were used to assess kindergartners' early literacy. Subtest 1 Letter-Word Identification and Subtest 13 Word Attack were used to assess early literacy with first graders.

Preliminary Longitudinal Findings: Kindergarten and First Grade Child Outcomes

In our presentation of the longitudinal findings, we will focus on child outcomes data gathered at three key points in time: the beginning of the kindergarten year, the end of the kindergarten year, and the end of the first grade year. At each point in time, we will present data on receptive vocabulary, mathematical skills, and early literacy. Data for children in Cohort 1, the older of our two cohorts, cover each of the three points in time. Data for children in Cohort 2 cover only the kindergarten year, as these children have not yet reached the end of first grade.

All analyses for this conventional, longitudinal approach compare children who attended ABC with those who did not attend a preschool program. These analyses were statistically controlled for ethnicity, gender, age, and children's free- and reduced-price lunch status. Analyses were conducted in STATA (StataCorp, 2005) using raw scores, and all estimated standard errors reflect clustering by classroom.

Child Outcomes at Kindergarten Entry

Our first set of analyses examines children's receptive vocabulary, mathematics, and early literacy skills at the beginning of kindergarten. Results from analyses using the conventional approach are summarized in Table 6 and are described in the following sections.

Table 6. Vocabulary, Mathematics, and Early Literacy Skills at Kindergarten Entry

	PPVT Receptive Vocabulary			
	ABC Score	No Pre-K Score	Difference	*p*
C1 K Entry	69.75	65.65	4.10	0.001**
C2 K Entry	68.76	65.94	2.82	0.021*
Pooled C1 & C2 K Entry	69.37	65.86	3.51	0.000**
	WJ-III Applied Problems Subtest			
	ABC Score	No Pre-K Score	Difference	*p*
C1 K Entry	15.52	15.34	0.18	0.560
C2 K Entry	15.46	14.65	0.81	0.013*
Pooled C1 & C2 K Entry	15.50	14.94	0.56	0.016*
	Pre-CTOPPP Print Awareness Subtest (Percentage Correct)			
	ABC Score	No Pre-K Score	Difference	*p*
C1 K Entry	79.48%	73.29%	6.19%	0.002**
C2 K Entry	76.74%	70.15%	6.59%	0.001***
Pooled C1 & C2 K Entry	78.52%	71.45%	7.07%	0.000***

Note: C1 represents Cohort 1 and C2 represents Cohort 2. K represents kindergarten. Raw scores are reported for the PPVT and WJ-III.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Receptive Vocabulary. For children in Cohort 1, the estimated effect of state-funded preschool on children's receptive vocabulary is statistically significant at the beginning of the kindergarten year. Children who attended the ABC preschool program at age 4 had PPVT scores that were 4.10 raw score points higher than children who did not attend pre-K. This represents an improvement of about 24% of the standard deviation for the control group.

For children in Cohort 2, the estimated effect of state-funded preschool on children's receptive vocabulary is also statistically significant at the beginning of the kindergarten year. Children who attended the ABC preschool program at age 4 had PPVT scores that were 2.82 raw score points higher than children who did not attend pre-K. This represents an improvement of about 15% of the standard deviation for the control group.

When we had comparable data from both cohorts at the same point in time, as we did with PPVT data at the beginning of kindergarten, we conducted additional analyses in which the samples from Cohorts 1 and 2 were combined, or pooled. This has the effect of creating larger ABC and No Pre-K samples, which allow for better estimates and the greater likelihood of having enough statistical power to detect the effects of preschool participation. Using the pooled dataset, the estimated effect of state funded pre-K on children's PPVT scores is statistically significant at the beginning of kindergarten. Children who attended the ABC program at age 4 had PPVT scores that were 3.51 raw score points higher than children who did not attend pre-K. This represents an improvement of about 20% of the standard deviation for the control group.

Mathematical Skills. At the beginning of the kindergarten year the estimated effect of state-funded preschool on Cohort 1 children's early math skills is not statistically significant for the ABC program. Children who attended the ABC preschool program as 4-year-olds scored 0.18 raw score points higher on the WJ-III Applied Problems subtest compared to children who did not attend pre-K. This represents an improvement of about 4% of the standard deviation for the control group.

At the beginning of the kindergarten year the estimated effect of state-funded preschool on Cohort 2 children's early math skills is statistically significant for the ABC program. Children who attended the ABC preschool program as 4-year-olds scored 0.81 raw score points higher on the WJ-III Applied Problems subtest compared to children who did not attend pre-K. This represents an improvement of about 19% of the standard deviation for the control group.

As we had WJ-III Applied Problems data for both cohorts at kindergarten entry, we again pooled data across the two cohorts. Using this dataset, the estimated effect of state pre-K on children's Applied Problems scores is statistically significant. Children who participated in ABC at age 4 scored 0.56 raw score points higher on the WJ-III Applied Problems subtest than children who did not participate in pre-K. This represents an improvement of about 13% of the standard deviation for the control group.

Early Literacy. For children in Cohort 1, the estimated effect of state-funded preschool on children's Print Awareness scores is statistically significant for the ABC preschool program at the beginning of the kindergarten year. Children who attended the ABC program answered 6.2%

more items correctly on the measure of print awareness than children who had not attended pre-K.

For children in Cohort 2, the estimated effect of state-funded preschool on children's Print Awareness scores is also statistically significant for the ABC preschool program at the beginning of the kindergarten year. Children who attended the ABC program answered 6.6% more items correctly on the measure of print awareness than children who had not attended pre-K.

When data are pooled across the two cohorts, the estimated effect of state pre-K on children's Print Awareness scores is statistically significant at the beginning of the kindergarten year. Children who participated in ABC at age 4 answered 7.1% more items correctly on the measure of print awareness than children who did not participate in pre-K.

Comparisons with Previous Data from the RDD. The longitudinal component of the ABC study allows for comparisons with the estimated effects presented in our earlier regression-discontinuity study (Hustedt et al., 2007) for children at kindergarten entry. The longitudinal methodology may introduce selection bias because some parents have chosen to send their children to the ABC program while others have not. In contrast, the RDD methodology corrects for selection bias in estimating the effects of ABC because all children included in the RDD sample participated in the ABC program.

The RDD study produced estimates of the ABC program's impacts at the start of kindergarten that were substantial on all three measures of learning: a broad test of verbal abilities, a math test, and a print awareness test. As shown in Table 7, the conventional longitudinal estimates were smaller than the RDD estimates and varied (but not in a consistent way) between the two cohorts studied. The good news is that despite the evident bias, the conventional estimates at kindergarten entry were statistically significant for all 3 measures using data pooled across the two cohorts, and for all 3 measures in 5 of 6 instances when data are analyzed separately for each of the cohorts. The bad news is that the bias appears to be substantial. This estimated bias is larger than in two other studies that compared RDD and conventional estimates for other preschool programs, likely reflecting the wider variation in the population attending ABC compared to the populations served in the other two studies (in Tulsa, Oklahoma and New Jersey's low-income urban districts).

Table 7 presents estimates of selection bias using data pooled across the two cohorts in the longitudinal study. These estimates range from a 35% reduction to a 69% reduction compared to the RDD. This much selection bias is a concern, but there are important additional considerations. First, the bias was largest for the Print Awareness measure, which will not be used in further follow-ups. Second, now that we know that bias is a substantial problem, we can collect better data on the family background characteristics for future analyses. Third, some fall-off in estimated effects on achievement occurs for all preschool programs, no matter how strong, so future estimated effects even if perfect would likely to be smaller than initial effects (though not necessarily smaller than effects by the time children reach first grade).

Table 7. RDD and Conventional Estimates of Effects for ABC at Kindergarten Entry

	Verbal Abilities	Math	Print Awareness (% Correct)
Regression-discontinuity	5.37*	1.22*	22.89%***
Conventional Cohort 1	4.10**	0.18	6.19%**
Conventional Cohort 2	2.82*	0.81*	6.59%***
Pooled Cohorts 1 & 2	3.51**	0.56*	7.07%***
Estimated Bias	35% too low	54% too low	69% too low

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

Child Outcomes at the End of Kindergarten

Another set of analyses examines children's receptive vocabulary, mathematics, and early literacy skills at the end of the kindergarten year. Findings from these analyses are described below.

Receptive Vocabulary. At the end of the kindergarten year, the estimated effect of state-funded preschool on children's PPVT scores approaches statistical significance for children in Cohort 1 ($p = .072$). Attending the ABC preschool program at age 4 is associated with an increase of 2.84 raw score points on the PPVT. The improvement is about 15% of the standard deviation for the control group.

At the end of the kindergarten year, the estimated effect of state-funded preschool on children's PPVT scores is not significant for children in Cohort 2. Attending the ABC preschool program at age 4 is associated with an increase of 0.88 raw score points on the PPVT. The improvement is about 6% of the standard deviation for the control group.

As we had PPVT data for both cohorts from the end of kindergarten, we again pooled data across the two cohorts. Due to the greater statistical power for analyses using this combined dataset, the estimated effect of state pre-K on children's receptive vocabulary scores was statistically significant. Children who participated in ABC at age 4 scored 2.15 raw score points higher on the PPVT than children who did not participate in pre-K. This represents an improvement of about 13% of the standard deviation for the control group.

A summary of the findings related to children's vocabulary skills at the end of kindergarten is provided in Table 8.

Table 8. Vocabulary Skills at the End of Kindergarten and First Grade

	PPVT Receptive Vocabulary			
	ABC Score	No Pre-K Score	Difference	*p*
C1 End of K	79.28	76.44	2.84	0.072^{+}
C2 End of K	80.16	79.28	0.88	0.484
Pooled C1 & C2 End of K	79.91	77.76	2.15	0.027*
C1 End of Grade 1	91.82	89.72	2.10	0.066^{+}

Note: C1 represents Cohort 1 and C2 represents Cohort 2. K represents kindergarten. Raw scores are reported.
$^{+}p < .10$. $^{*}p < .05$.

Mathematical Skills. At the end of the kindergarten year the estimated effect of the ABC preschool program on children's early math skills is not statistically significant for children in Cohort 1. Attending the ABC preschool program at age 4 is associated with an increase of 0.60 raw score points on the WJ-III Applied Problems subtest. This represents an improvement of about 16% of the standard deviation for the control group.

At the end of the kindergarten year the estimated effect of the ABC preschool program on children's early math skills as measured by WJ-III Applied Problems subtest scores also is not statistically significant for children in Cohort 2. Attending the ABC preschool program at age 4 is associated with a decrease of 0.06 raw score points on the WJ-III Applied Problems subtest. This represents a decrease of about 2% of the standard deviation for the control group.

As we had WJ-III Applied Problems data for both cohorts from the end of kindergarten, we again pooled data across the two cohorts. Using this dataset, the estimated effect of state pre-K on children's Applied Problems scores is not statistically significant. Children who participated in ABC at age 4 scored 0.36 raw score points higher on the WJ-III Applied Problems subtest than children who did not participate in pre-K. This represents an improvement of about 9% of the standard deviation for the control group.

A summary of the findings related to children's mathematics skills at the end of the kindergarten year is provided in Table 9.

Table 9. Mathematical Skills at the End of Kindergarten and First Grade

	WJ-III Applied Problems Subtest			
	ABC Score	No Pre-K Score	Difference	*p*
C1 End of K	20.10	19.50	0.60	0.129
C2 End of K	19.83	19.89	-0.06	0.835
Pooled C1 & C2 End of K	20.03	19.67	0.36	0.158
C1 End of Grade 1	25.29	24.62	0.67	0.054^{+}
	WJ-III Calculation Subtest			
	ABC Score	No Pre-K Score	Difference	*p*
C1 End of Grade 1	9.12	8.72	0.40	0.045*
	WJ-III Math Fluency Subtest			
	ABC Score	No Pre-K Score	Difference	*p*
C1 End of Grade 1	27.52	26.23	1.29	0.197

Note: C1 represents Cohort 1 and C2 represents Cohort 2. K represents kindergarten. Raw scores are reported.
$^{+}p < .10$. $^{*}p < .05$.

Early Literacy. At the end of kindergarten, the effect of state-funded preschool on children's Print Awareness scores is not statistically significant for the ABC program participants in Cohort 1.

Beginning with the Spring 2007 child assessment period, we adapted our assessment battery to include early literacy measures that were more appropriate for children in kindergarten

and first grade. For children in Cohort 2, who were completing kindergarten in Spring 2007, the effect of state-funded preschool on children's Letter-Word Identification scores is not statistically significant for ABC program participants. The effect of state-funded preschool on children's Sound Awareness scores (substitution, deletion, and rhyming components) also is not statistically significant for Cohort 2 ABC participants at the end of kindergarten. Table 10 summarizes the findings related to children's early literacy skills at the end of the kindergarten year.

Table 10. Early Literacy Skills at the End of Kindergarten and First Grade

	Pre-CTOPPP Print Awareness Subtest (Percentage Correct)			
	ABC Score	No Pre-K Score	Difference	*p*
C1 End of K	95.35%	95.92%	-0.57%	0.537
	WJ-III Letter-Word Identification Subtest			
	ABC Score	No Pre-K Score	Difference	*p*
C2 End of K	22.48	22.19	0.29	0.598
C1 End of Grade 1	34.97	33.17	1.80	0.013*
	WJ-III Word Attack Subtest			
	ABC Score	No Pre-K Score	Difference	*p*
C1 End of Grade 1	14.39	14.23	0.16	0.803

Note: C1 represents Cohort 1 and C2 represents Cohort 2. K represents kindergarten. Raw scores are reported for the WJ-III.
* $p < .05$.

Child Outcomes at the End of First Grade

A third set of analyses from our longitudinal study looks at children's receptive vocabulary, mathematics, and early literacy skills at the end of first grade. Conventional estimates from these analyses are described below.

Receptive Vocabulary. At the end of first grade, the estimated effect of state-funded preschool on children's PPVT scores approaches statistical significance for children in Cohort 1 ($p = .066$). Attending the ABC preschool program at age 4 is associated with an increase of 2.10 raw score points on the PPVT. The improvement is about 13% of the standard deviation for the control group. Findings related to first graders' receptive vocabulary skills are shown in Table 8.

Mathematical Skills. By the end of first grade the estimated effect of the ABC preschool program on children's Applied Problems subtest scores approaches statistical significance for children in Cohort 1. Attending the ABC preschool program at age 4 is associated with an increase of 0.67 raw score points on the Applied Problems subtest, representing an improvement of about 15% of the standard deviation for the control group.

First graders also were assessed using the Calculation and Math Fluency subtests of the WJ-III, in order to gather better information about mathematics skills in slightly older children. The estimated effect of the ABC program on children's Calculation subtest scores is statistically significant for children in Cohort 1. Attending the ABC program at age 4 is associated with an increase of 0.40 raw score points on the Calculation subtest, representing an improvement of about 17% of the standard deviation for the control group. The estimated effect of the ABC program on children's Math Fluency subtest scores is not statistically significant for children in Cohort 1. Attending the ABC program at age 4 is associated with an increase of 1.29 raw score points on the Math Fluency subtest, representing an improvement of about 12% of the standard deviation for the control group.

As the Calculation, Math Fluency, and Applied Problems subtests comprise the Broad Math Battery of the WJ-III, we also conducted analyses of children's Broad Math scores. The estimated effect of the ABC program on children's Broad Math scores is statistically significant for children in Cohort 1 ($p = .022$). Attending the ABC program at age 4 is associated with an increase of 3.40 raw score points on the Broad Math Battery. Findings related to children's mathematics skills at the end of first grade are summarized in Table 9.

Early Literacy. For children in Cohort 1, who were completing first grade during Spring 2007, the effect of state-funded preschool on children's Letter-Word Identification scores is statistically significant for ABC program participants. Attending the ABC preschool program at age 4 is associated with an increase of 1.80 raw score points on the Letter-Word Identification subtest, representing an improvement of about 20% of the standard deviation for the control group. The effect of state-funded preschool on children's Word Attack scores is not statistically significant for Cohort 1 ABC participants at the end of first grade.

As the Letter-Word Identification and Word Attack subtests comprise the Basic Reading Skills Battery of the WJ-III, we conducted additional analyses of children's scores on Basic

Reading Skills. The estimated effect of the ABC program on children's Basic Reading scores approaches statistical significance for children in Cohort 1 at the end of first grade ($p = .069$). Attending the ABC program at age 4 is associated with an increase of 2.45 raw score points on the Basic Reading battery. Table 10 presents the findings related to children's early literacy skills at the end of first grade.

Adjusting for Selection Bias: Re-examining Results from the End of Kindergarten and First Grade

As the ABC study includes an RDD component in addition to the conventional, longitudinal component, it is possible to adjust results gathered through the conventional approach based on findings from the RDD approach. By comparing results from both approaches at kindergarten entry – the child assessment point common to both the RDD and conventional designs – we can measure of the magnitude of selection bias present in our conventional results. The conventional estimates are likely to underestimate the true effects of the ABC program, since unlike the RDD results they are subject to selection bias. Indeed, we found the conventional estimates to be about 50% lower than the RDD estimates from the beginning of kindergarten (see Table 7). In Table 11, we present a final set of results from the end of kindergarten and the end of the first grade year. These results have been mathematically adjusted for the magnitude of selection bias we found to be present at the beginning of kindergarten. 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.

Summary and Discussion

The results presented in this report provide further documentation of the positive impacts of the Arkansas Better Chance Program. An earlier analysis (Hustedt et al., 2007) capitalized upon a sophisticated regression-discontinuity research design, showing that ABC participants had higher vocabulary, mathematics, and print awareness scores when entering kindergarten, compared to children who did not attend pre-K. Our first longitudinal report extends those findings by focusing on a larger sample of children as they progressed through kindergarten and first grade, and also provides information about classroom quality.

Classroom Quality in the Arkansas Better Chance Program

Findings from structured classroom observations conducted during the first year of the longitudinal study allow us to examine overall quality in the ABC classrooms. As classroom quality is likely to be a key factor through which state-funded ABC prekindergarten programs produce positive effects for children, gathering data on the classroom environment was a critical early step in this research. Our results show that the quality of ABC prekindergarten classrooms was generally high. The mean overall classroom quality score on the ECERS-R (Harms et al., 2005) was 5.26, using a scale that ranges from 1 to 7. As a point of reference, the test developers characterize a score of 5 as "good".

Table 11. Child Outcomes at the End of Kindergarten and First Grade, Adjusted for Selection Bias

	Point Difference Between ABC and No Pre-K, Unadjusted	Point Difference Between ABC and No Pre-K, Adjusted for Selection Bias
PPVT Receptive Vocabulary		
C1 End of K	2.84	4.35
C2 End of K	0.88	1.35
C1 End of Grade 1	2.10	3.21
WJ-III Applied Problems Subtest		
C1 End of K	0.60	1.31
C2 End of K	-0.06	-0.13
C1 End of Grade 1	0.67	1.46
WJ-III Calculation Subtest		
C1 End of Grade 1	0.40	0.87
WJ-III Math Fluency Subtest		
C1 End of Grade 1	1.29	2.81
Pre-CTOPPP Print Awareness Subtest (Percentage Correct)		
C1 End of K	-0.57%	-1.85%
WJ-III Letter-Word Identification Subtest		
C2 End of K	0.29	0.44
C1 End of Grade 1	1.80	2.75
Word Attack Subtest (WJ-III Raw Scores)		
C1 End of Grade 1	0.16	0.24

Note: C1 represents Cohort 1 and C2 represents Cohort 2. K represents kindergarten. Raw score points are reported for the PPVT and WJ-III.

The mean overall score for the SELA (Smith et al., 2001), which measures classroom support of literacy skills, was 3.28 on a scale that ranges from 1 to 5. This is a positive result, indicating that, on average, ABC classrooms are providing a good early literacy environment for the children who are enrolled. Finally, the mean overall score on the PCMI (Frede et al., 2005), which measures the early mathematics environment, was 2.37 on a scale from 1 to 5. This is a less positive result. Though there was wide variability in PCMI item scores across different ABC classrooms, the average ABC classroom was providing limited support for children's early mathematics skills. The PCMI items with the highest mean scores were the four items dealing with the types of classroom materials available to support mathematics skills, while the lowest scoring items involved teaching behaviors. These findings suggest that while some improvement in the early mathematics environment could be gained by providing additional or better classroom materials, children may benefit the most from efforts aimed at increasing teachers' focus on mathematics in the prekindergarten classroom. For example, teacher professional development activities could be used to emphasize the importance of early mathematics learning and to provide age-appropriate strategies for incorporating better mathematics teaching into ABC classrooms.

The overall classroom quality findings for the ABC program show a pattern similar to those from NIEER's study of the Abbott Preschool Program in New Jersey (Frede et al., 2007). In that study, Frede and colleagues found that overall classroom quality and classroom support for early literacy were good, but that classroom support for early mathematics was, on average, lacking. The relatively lower scores on classroom mathematics environments suggest that this may be an area in need of greater emphasis in state-funded pre-K more generally, although further research is needed.

Child Outcomes in Kindergarten and First Grade

Turning toward the outcomes associated with ABC participation as children progress through school, we examined children's language, literacy, and math scores in both kindergarten and first grade. At this point in our longitudinal study, kindergarten data for two cohorts of children are available, along with first grade data for Cohort 1, our oldest cohort of children.

Findings from the beginning of kindergarten were similar for children in Cohorts 1 and 2. At kindergarten entry, former ABC pre-K participants in both cohorts scored significantly higher on assessments of vocabulary and print awareness compared to children who had not attended pre-K. For Cohort 1 there were no significant differences on the mathematics measure when comparing kindergartners who had attended ABC with those who had not attended pre-K. However, in Cohort 2, and when data from Cohorts 1 and 2 were pooled, ABC participants scored significantly higher on our assessment of early mathematics skills than children who did not attend pre-K.

Findings from the end of kindergarten were also similar for children in Cohorts 1 and 2. At the end of kindergarten, there were no differences on the mathematics or early literacy measures for children based on ABC participation. There was a trend for ABC participants in Cohort 1 to score higher on the vocabulary measure than children who did not attend pre-K, though the difference between the two groups did not reach statistical significance. While no

such relationship was found for children in Cohort 2, when data were pooled across the two cohorts, children who attended ABC scored significantly higher on the vocabulary measure than children who had not attended pre-K.

At the end of first grade, there was a trend for ABC participants in Cohort 1 to score higher on the vocabulary measure than children who had not participated in pre-K, though the difference between the two groups' vocabulary scores did not reach statistical significance. We will re-examine this finding by pooling data across the two cohorts once first grade data are available for Cohort 2.

At the end of first grade we found evidence for a continuing effect of ABC participation on children's early mathematics skills. ABC participants scored higher on the Calculation measure than children who had not participated in pre-K. There was also a trend for ABC participants to score higher on the Applied Problems measure, although this result was not statistically significant. No differences were found between the two groups of children on the Math Fluency measure. However, there was a statistically significant difference on the Broad Math Battery from the Woodcock-Johnson III, which uses results from the three subtests mentioned above: Calculation, Math Fluency, and Applied Problems. Overall, children who had participated in ABC tended to have higher scores on the Broad Math Battery than children who had not participated in pre-K.

Finally, we found evidence for a continuing effect of ABC participation on children's literacy skills at the end of first grade. Children who attended the ABC program scored higher on the Letter-Word Identification subtest than children who had not attended pre-K. There was no difference in children's scores on the Word Attack measure based on participating in ABC. However, the results from an analysis of Basic Reading scores, which are based on the Letter-Word Identification and Word Attack subtests, show that there was a non-significant trend for ABC participants to score higher than children who did not attend pre-K. Again, we report end-of-first grade results only for Cohort 1, as children in Cohort 2 have not yet completed first grade.

While the current results point toward continuing benefits for children who participated in the ABC program, the results are not as consistently strong as those outlined in our earlier regression-discontinuity report (Hustedt et al, 2007). In general, though, the RDD results from Arkansas are comparable to those from a number of other states in which similar studies of state-funded prekindergarten have been conducted (Wong et al., 2008). This similarity increases our confidence that RDD findings provide unbiased estimates of the effects of the ABC program at the beginning of kindergarten. While the RDD methodology allows for correction of selection bias at kindergarten entry, unfortunately it cannot provide an estimate of effects at later points in time. If we employed it a year later, it would provide an estimate of the added effects of kindergarten. Thus, we employed a second approach to obtain estimates beyond kindergarten entry using a conventional no-treatment comparison group, but employ it at kindergarten entry as well so we can compare its estimates to the RDD estimates and determine the extent of any selection bias.

Results from the conventional analyses presented in this report do provide substantial evidence of selection bias in the longitudinal sample. We used the RDD component of the study to estimate the magnitude of selection bias in the longitudinal analyses, and adjusted the conventional, longitudinal results accordingly. For measures that are comparable across the assessment periods, these revised estimates of the effects of ABC on language, mathematics, and early literacy skills are fairly consistent with those presented in our earlier report on RDD findings. The statistically significant positive effects in mathematics and literacy measures at the end of first grade should be considered extremely positive findings, considering that statistical significance is based on the unadjusted results that underestimate the effects of ABC participation.

The results in this report may also underestimate the effects of ABC participation due to some limitations in the early longitudinal data that are being addressed in future rounds of data collection. One general concern is that there may be differences in family income between ABC participants and children who did not attend pre-K. 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 groups of children who attended ABC and those not attending pre-K were not entirely comparable at the beginning of the kindergarten year. 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 state public school database (APSCN) in early 2007. The analyses described in this report were controlled for children's free- and reduced-price lunch status, but data on children's lunch status were not available for some children, who could not be matched to the database. Now that we know that selection bias is a substantial problem, for future analyses we will collect more data on family background characteristics. This information can be used to statistically adjust the conventional longitudinal estimates in an effort to reduce selection bias. However, it must be acknowledged that there is no way to judge the success of this effort in advance.

A limitation of the early literacy data gathered for kindergartners is that since the Pre-CTOPPP is primarily intended for use with prekindergarten populations, most children provide correct answers for very high percentages of the items by the end of the kindergarten year. Although ABC participants from Cohort 1 correctly answered more Print Awareness items at the beginning of kindergarten, differences between the ABC and No Pre-K groups may have disappeared by the end of the year because of a ceiling effect. In fact, children in both the ABC and No Pre-K groups answered more than 95% of the Print Awareness items correctly by the end of kindergarten. Essentially, the skills tested by the Pre-CTOPPP may be present in nearly all children who have completed kindergarten, regardless of any continued effect of the ABC program.

As children in the longitudinal study get older, the tests must constantly be changed in order to test the appropriate grade level material. In response to our concern that the initial assessment battery was not well enough aligned with children's developing literacy and math skills during the early years of school, we began updating our assessment battery beginning with the Spring 2007 child assessment period. At that time, we started phasing in more advanced assessments of early literacy and mathematics. These assessments are designed for use with

slightly older children, and should allow us to better track the development of children's early literacy skills in future years of the study. The early findings from our new assessment battery are encouraging, with positive and statistically significant impacts evident at the end of first grade on measures of both early literacy and math.

However, even with our new assessment battery in place, the early literacy findings from the end of kindergarten are more difficult to interpret. The absence of a relationship between ABC attendance and early literacy scores at the end of kindergarten is puzzling, given that such relationships exist both at the beginning of kindergarten and at the end of first grade. Analyses of additional rounds of first grade (Cohort 2) and second grade (Cohort 1) data to be gathered during Spring 2008 should help to clarify this relationship.

The longitudinal nature of the Arkansas Better Chance evaluation has allowed us to begin incorporating the methodological refinements described above as we continue tracking the estimated effects of the ABC program over time. Early longitudinal results show that the benefits of participating in the ABC program remain evident at the end of first grade, as demonstrated through children's language, literacy, and mathematics skills. During the current 2007-2008 data collection cycle, the two cohorts of children in the longitudinal study are in first and second grades, respectively. First grade data from Cohort 2 in this cycle will be added to the first grade Cohort 1 data from the previous school year. This will allow for pooled analyses of the first grade results that have greater statistical power. Second grade data from this cycle will allow for new estimates of the effects of ABC prekindergarten by the end of grade 2. All children will continue to participate in the study through the 2009-2010 school year, as they progress through the next two grades of elementary school. Future reports will provide revised estimates of the impacts of participating in the ABC program as children get older.

References

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 (PPVT-III).* 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.

Frede, E., Lamy, C. E., & Boyd, J. S. (2005) *Not just calendars and counting blocks: Using the NAEYC/NCTM joint position statement "Early childhood mathematics: Promoting good beginnings" as a basis for measuring classroom teaching practices and their relationship to child outcome.* Paper presented at the annual National Association for the Education of Young Children conference, Washington, DC.

Frede, E., Weber, M., Hornbeck, A., Stevenson Boyd, J., & Colon, A. (2005). *Preschool Classroom Mathematics Inventory*. Available from the first author at efrede@nieer.org.

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.

Harms, T., Clifford, R., & Cryer, D. (2005). *Early Childhood Environment Rating Scale Revised Edition (ECERS-R)*. New York: Teachers College Press.

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.

Smith, S., Davidson, S., & Weisenfeld, G. (2001). *Supports for Early Literacy Assessment for early childhood programs serving preschool-age children*. New York: New York University.

StataCorp. (2005). *Stata Statistical Software: Release 9.* College Station, TX: StataCorp LP.

Trochim, W. M. K. (1984). *Research design for program evaluation*. Beverly Hills, CA: Sage.

Wagner, R., Torgeson, J., & Rashotte, C. (1999). *Comprehensive Test of Phonological Processing (CTOPP).* Austin, TX: Pro-Ed.

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: Applying the RDD Approach to the ABC Study

The regression-discontinuity design (RDD) takes advantage of a strict enrollment policy that determines enrollment using the child's date of birth to define the groups. By relying on this assignment rule, one that is unlikely to be related to child and family characteristics, the RDD seeks to reduce the likelihood of selection bias. Thus, rather than compare children who attended and did not attend the program (raising concerns that the same child and family factors that led program eligibility or a family seeking to enroll a child in the program also contribute to differences in learning and development), the RDD approach compares two groups of children who enroll in the ABC program. One group has completed the program and the other is just entering.

One way to interpret the RDD approach is to view it as similar to a randomized trial for children near the age cutoff. The RDD creates groups that *at the margin* differ only in that some were born a few days before the age cutoff and others a few days after the cutoff. When these children are about to turn 5 years old, the slightly younger children will enter the preschool program and the slightly older children will enter kindergarten having already attended the ABC program. If all of the children are tested at that time, the difference in their scores can provide an unbiased estimate of the preschool program's effect under reasonable circumstances. Of course, if only children with birthdays only a few days on either side of the age cutoff were included in a study, the sample size would be unreasonably small.

Alternatively, the RDD can be viewed as modeling the relationship between an assignment variable (age) and measures of children's learning and development. The pre-cutoff sample is used to model the relationship prior to treatment. The post-cutoff sample is used to model the relationship after the treatment. This approach can be applied to wider age ranges around the cutoff. However, its validity depends on correctly modeling the relationship. As there is always some uncertainty about what this looks like (is it linear, and if not what does the curve look like?), we test out a variety of different models (different functional forms for the equation) to see what fits the data best, in addition to conducting other tests of the RDD assumptions. Under either view, it is important that there is minimal misallocation (exceptions to the rule) around the age cutoff.

To identify the proper functional form for our RDD analyses, we conducted a graphical analysis and a series of parametric regressions with alternate specifications.

We begin with simple graphs of each outcome. As shown in Figures 1-3, two types of lines are fitted onto the scatterplots on each side of the cutoffs. Plot 1 depicts a linear regression line, and Plot 2 shows a non-parametric regression line based on locally weighted scatterplot smoothing, called Lowess. This strategy is often used for data exploration because it relaxes assumptions about the form of the relationship between the assignment and outcome (Cleveland & Devlin, 1988). For each y_i, a smoothed value is obtained by weighted regressions involving only those observations within a local interval. Observations closer to y_i are weighted more heavily than those farther away.

We next run a series of regressions to obtain parametric estimates of the treatment effect. To describe the causal relationship of state pre-K participation on children's achievement scores we model the latter. For the ith individual in classroom j, we write:

$$Y_{ij} = a + BX_{ij} + \beta_1(\text{Pre-K})_{ij} + g(AV)_{ij} + \varepsilon_i$$

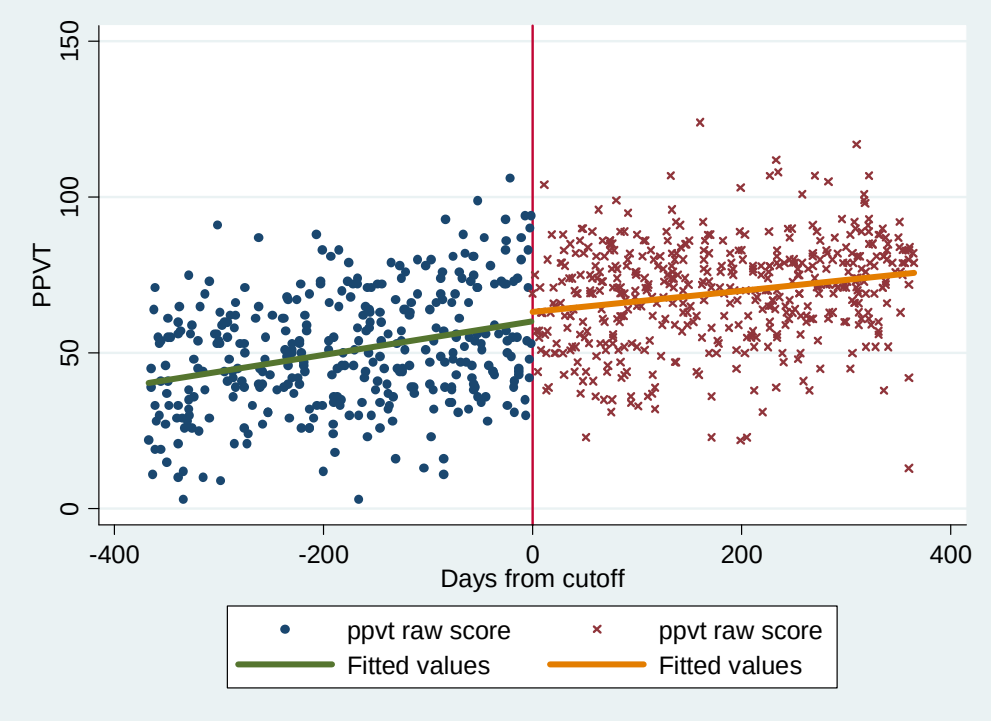
where Y_{ij} is student i's outcome, X_{ij} is a vector of student characteristics including gender, race/ethnicity. Pre-K_{ij} is a dichotomous indicator variable such that T=1 for treatment and T=0 for no treatment, and $g(AV)_{ij}$ is a smooth function of the continuous assignment variable. We check for robustness of our estimates by considering a number of alternative specifications for $g(AV)_{ij}$, including polynomials and interaction terms. The order of the polynomial approximation to the $g(AV)_{ij}$ function is determined by examining the statistical significance of the higher order and interaction terms. Following Trochim (1984), when the functional form of the regression model is ambiguous, we overfit the model by including more polynomial and interaction terms than needed, yielding unbiased but less efficient estimates. In all the parametric analyses we use Huber-White standard errors adjusted for clustered data at the classroom level. Table A1 provides impact estimates using linear, quadratic, and cubic models.

As a final parametric check on functional form, we truncate the dataset to include only observations near the cutoff. In placing greater weight on these we eliminate the influence of extreme assignment variable values that often play a disproportionate role in mis-specifying functional form. So we rerun the parametric analyses including only those children with birthdates within three and six months each side of the cutoff.

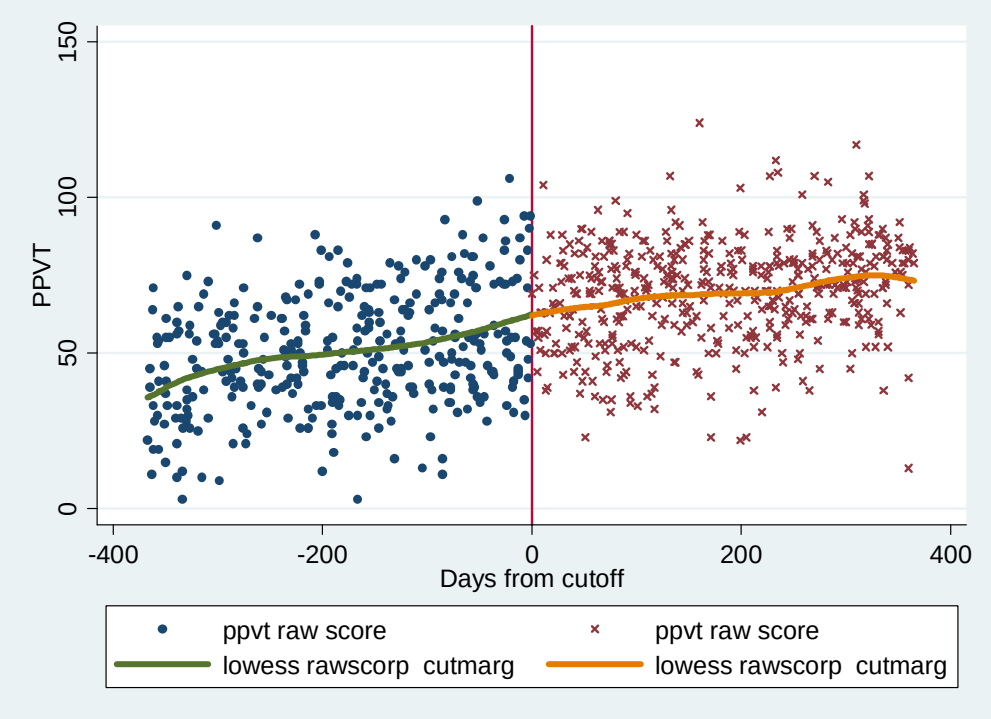
Graphical and parametric analyses provide evidence that the response function was linear for PPVT, math, and print awareness outcomes. Estimates were robust with respect to narrowing the time window around the birthdate cutoff to 6 and 3 months. We choose the linear specification because of evidence from graphical plots (Figures 1, 2, and 3) and lack of statistical significant in the higher order terms.

One issue with the estimates is that the PPVT and mathematics results are somewhat sensitive to *functional form* specification and so we view these estimates with uncertainty. For print awareness, the estimated effect was highly consistent across all models. For PPVT and mathematics results, the quadratic form yielded similar estimates but the cubic form did not. However, neither quadratic nor cubic models were good fits to the data. Furthermore, estimates from a linear model with only a 3-month interval around the birthdate cutoff yielded estimates that were highly similar those found for the full sample.

Figure 1. Linear and Lowess Plots of PPVT Receptive Vocabulary Results

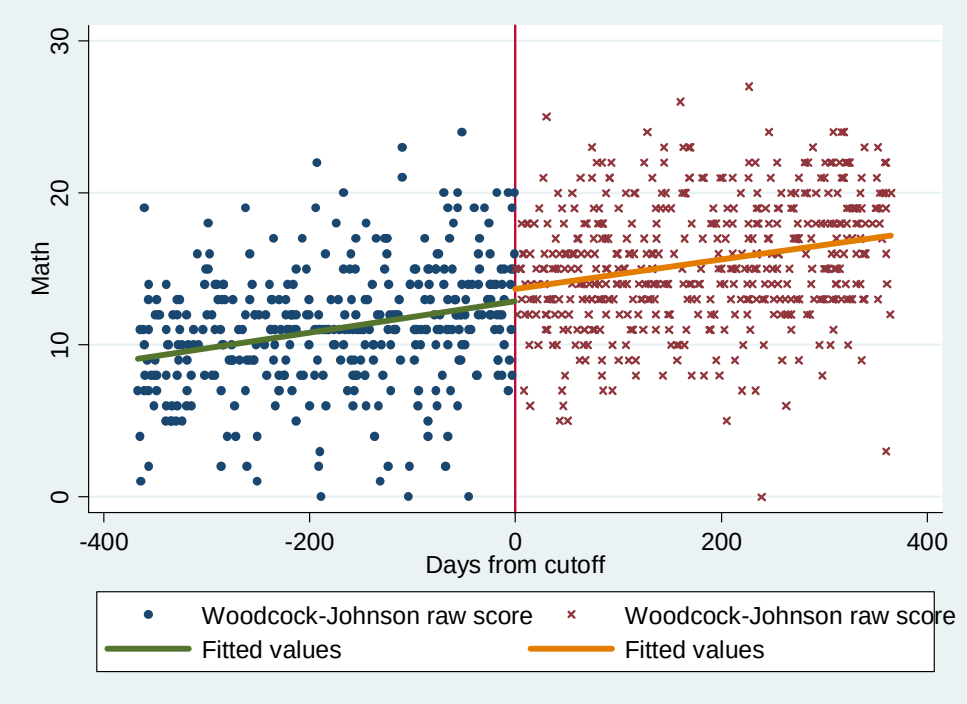


(1) Linear plot

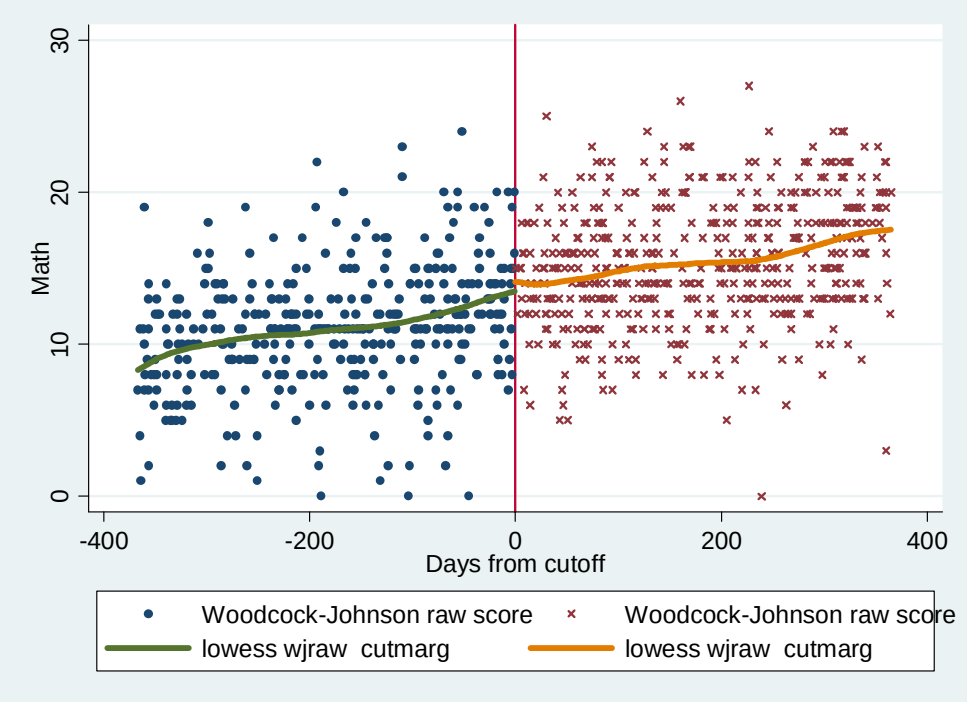


(2) Lowess Plot

Figure 2. Linear and Lowess Plots of WJ-III Applied Problems Results

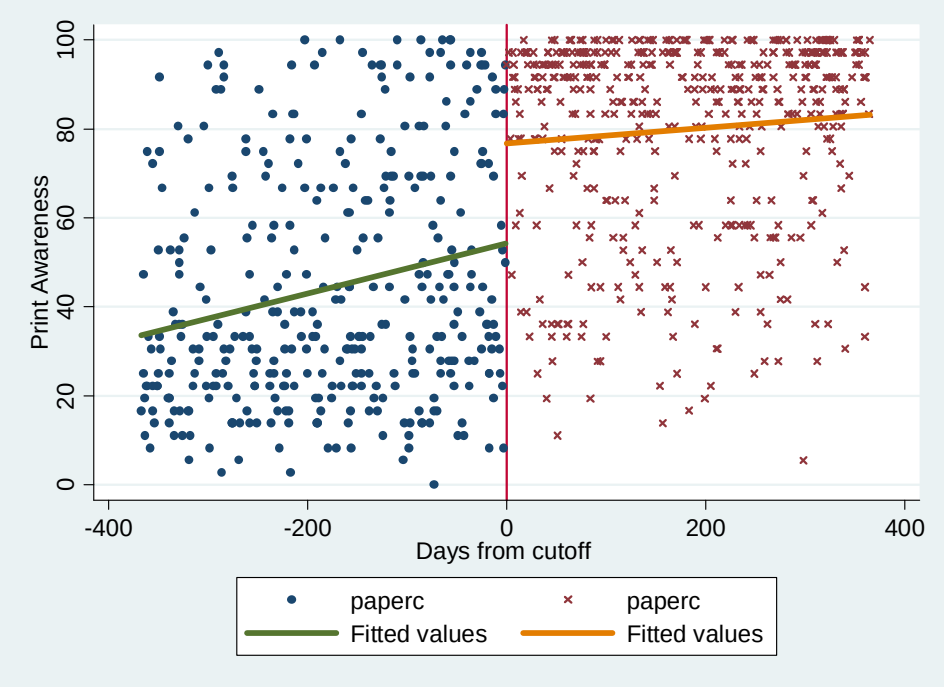


(1) Linear plot

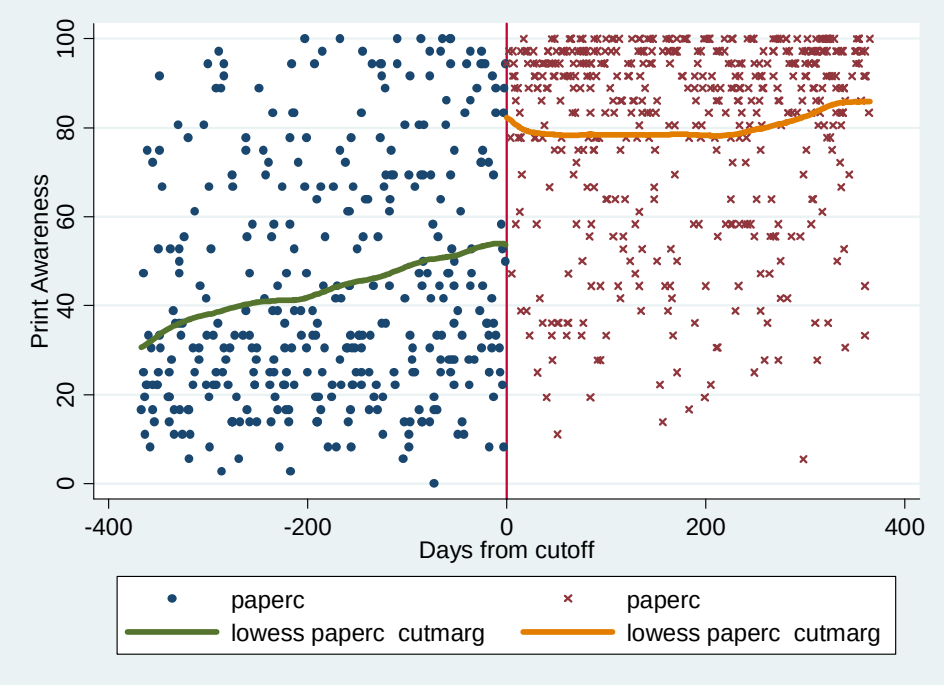


(2) Lowess plot

Figure 3. Linear and Lowess Plots for Pre-CTOPPP Print Awareness Results



(1) Linear Plot



(2) Lowess Plot

Table A1. Impact Estimates Using Linear, Quadratic, and Cubic Models

	Empirically Identified Functional Form	Parametric Models Used in Analysis				
		Linear	Quadratic	Cubic	Truncated at 3 Months	Truncated at 6 Months
PPVT	Linear	5.37*	5.35	1.37	5.71	3.44
		(2.30)	(3.57)	(4.63)	(4.61)	(3.39)
Math	Linear	1.22*	1.30+	0.08	.82	.66
		(.53)	(.74)	(.97)	(.93)	(.73)
Print Awareness	Linear	22.89*	27.57*	23.96*	26.92*	24.58*
		(3.90)	(5.44)	(7.39)	(7.31)	(5.53)

Robust standard errors in parentheses
$^{+}p < .10$. $^{*}p < .05$. None of the higher order terms and interaction terms was significant at $p <. 05$.

In the cubic model for PPVT, the quadratic and cubic terms approached significance ($p = .09$, $p = .08$, respectively). In the cubic model for Math, the quadratic term and quadratic interaction term approached significance ($p = .10$, $p = .08$, respectively).