Continued Impacts of New Mexico PreK on Children's Readiness for Kindergarten:

Results from the Third Year of Implementation

Jason T. Hustedt, Ph.D.
W. Steven Barnett, Ph.D.
Kwanghee Jung, Ph.D.
Alexandra Figueras-Daniel, M.A.

National Institute for Early Education Research
Rutgers University, Graduate School of Education

September 2009

Acknowledgements

Continued Impacts of New Mexico PreK
on Children's Readiness for Kindergarten:
Results from the Third Year of Implementation

Executive Summary

This multi-year evaluation of New Mexico's state-funded PreK initiative uses a rigorous research design to estimate the impacts of participating in New Mexico PreK at age 4. The current report focuses on the initiative's third year of operation to assess the academic skills of kindergartners who attended New Mexico PreK during the 2007-2008 school year. Children's receptive vocabulary, math, and early literacy skills were measured using a sample of 1,333 PreK participants from across the state. Our results show that New Mexico PreK continues to produce positive impacts on children's early learning when they start kindergarten. These results are generally consistent with findings from similar studies focusing on the first and second years of the initiative.

More specifically:

1. Children who attended New Mexico PreK during the 2007-2008 school year scored higher on assessments of early math and literacy skills in comparison to children who did not attend. These skills include addition and subtraction, telling time, knowledge of letters, and familiarity with words and book concepts. Gains in early math and literacy at kindergarten entry can be attributed to participating in New Mexico PreK programs the previous year.

2. Separate sets of analyses conducted for PreK programs offered by the New Mexico Public Education Department (PED) and the Children, Youth and Families Department (CYFD) show that PED and CYFD PreK programs produce similar results.

3. When we combine child assessment data from the first three years of the New Mexico PreK program, we find further evidence that New Mexico PreK produces positive impacts on children's early math, language, and literacy skills.

Additional data on the impacts of New Mexico PreK will be gathered as a second phase of this evaluation begins in fall 2009. This study will also include a new component, investigating longer-term impacts of participating in New Mexico PreK as its first cohort of participants reaches third grade.

Introduction

This is the third annual report on the effects of the New Mexico PreK initiative on children's language, literacy, and math skills when they start kindergarten. These reports summarize results from a multi-year evaluation of New Mexico PreK by the National Institute for Early Education Research (NIEER) at Rutgers University that started during the 2005-2006 school year. A primary goal of our ongoing evaluation is to examine the impacts of the statewide PreK initiative for 4-year-olds as it undergoes a planned expansion. An additional goal is to provide separate estimates for PreK classrooms operated by the state Children, Youth and Families Department (CYFD) and the state Public Education Department (PED). This report is organized into four main sections:

1. Background information about state-funded prekindergarten programs nationwide and how New Mexico PreK fits into this broader context;
2. Methodological details about our research, including a discussion of our sample and the instruments used to gauge children's school readiness in New Mexico;
3. Findings from our most recent statistical analyses, including estimates of the impacts of the statewide New Mexico PreK program as well as separate estimates for PED and CYFD sites; and
4. A discussion of the results and their implications.

Background

School enrollment of 4-year-olds has increased tremendously since the 1960s. As a result, most children in the U.S. now have their first school experiences in preschool classrooms even before entering kindergarten (Barnett & Yarosz, 2007). State-funded pre-K programs have played a major role in the expansion of educational opportunities for young children, especially during the past two decades (Barnett, Epstein, Friedman, Stevenson Boyd, & Hustedt, 2008; Mitchell, 2001). By 2008, state pre-K programs were operating in 38 states, and served more than 1.1 million children nationwide using $4.6 billion in state dollars (Barnett et al., 2008). At age 4, one in four American children now attends a program that can be classified as state pre-K. Some states have committed to making state pre-K available to all 4-year-olds whose parents would like them to attend: Florida, Georgia, Illinois, Iowa, New York, Oklahoma, and West Virginia (Barnett, Friedman, Hustedt, & Stevenson Boyd, 2009). To date, Oklahoma has come closest to meeting this goal, with more than 70% of the state's 4-year-olds currently attending state-funded pre-K programs during the year before kindergarten (Barnett et al., 2008).

A common emphasis of state pre-K initiatives across the U.S. is to prepare children for kindergarten, recognizing that effective early childhood programs help them develop the knowledge, skills, and dispositions needed for school success, such as rich vocabulary and complex sentence structure, self-regulation, and cooperative play. Yet, state policies regarding the availability and areas of emphasis of prekindergarten initiatives tend to vary widely from state to state.

Also, despite many years of overall growth in state pre-K initiatives, there have been few research studies evaluating the statewide effectiveness of these programs. Among the

evaluations that do exist, even fewer are methodologically rigorous (Gilliam & Zigler, 2004). Previous research with model preschool initiatives shows that high-quality and well-funded programs can make valuable contributions to children's learning and development (Barnett, 2002). For example, studies of well-known initiatives including the High/Scope Perry Preschool program, the Abecedarian Early Childhood Intervention program, and the Chicago Child-Parent Centers show that these types of programs produce economic benefits that greatly outweigh their costs (Barnett, 1996; Masse & Barnett, 2002; Reynolds, Temple, Robertson, & Mann, 2002). Benefits include higher achievement test scores and lower rates of special education placements and grade repetition, as well as long-term effects such as improved high school graduation rates and reduced crime and delinquency rates. State pre-K programs typically are larger scale initiatives than many of the model programs that have been intensively studied, and are not as well funded. However, a number of states now offer prekindergarten programs that are both high in quality and widely available. Results from a widely cited and methodologically rigorous study of state pre-K participants in the city of Tulsa, Oklahoma (Gormley, Gayer, Phillips, & Dawson, 2005; Gormley, Phillips, & Gayer, 2008) show that high-quality state pre-K programs can generate statistically significant short-term impacts on a range of academic content areas, for children across a range of racial and ethnic groups, and for children from both low- and middle-income backgrounds.

The New Mexico PreK initiative began during the 2005-2006 school year and offers voluntary center-based prekindergarten to 4-year-olds across New Mexico. Funds to operate a New Mexico PreK classroom are provided by the state through CYFD and PED, and in order to receive these funds, sites must submit funding proposals that are "… evaluated on the percentage and number of public elementary schools in the community that are not meeting the proficiency component required for calculating adequate yearly progress and that are serving children, at least sixty-six percent of whom live within the attendance zone of a Title 1 elementary school" (Pre-Kindergarten Act, NMSA 1978 § 32A-23-6, 2005). Additional criteria used by the state to prioritize proposed sites for funding include the number of 4-year-olds in the community that will be served and the adequacy of prekindergarten sites already operating in that community.

New Mexico PreK classrooms have maximum class sizes of 20 with staff-child ratios of 1:10, and offer a variety of comprehensive and family support services in addition to the emphasis on early childhood education. Standards requiring all lead teachers to have bachelor's degrees and licensure in early childhood education are being phased in over time (Barnett et al., 2008), and it is expected that lead teachers who have worked in New Mexico PreK for 5 years will have bachelor's degrees and early childhood licensure by fall 2010. Participating providers include public schools as well as private providers such as Head Start, child care facilities, faith-based centers, and tribal programs. PreK program sites are administered by either PED or CYFD, depending on which agency provided funding, and overall administrative responsibility for the statewide PreK initiative is shared by PED and CYFD.

The Western and Northern Plains states have tended to lag behind other parts of the U.S. in making state-funded prekindergarten programs available to young children (Barnett, Hustedt, Hawkinson, & Robin, 2006). However, New Mexico's PreK initiative has grown quickly since it began in fall 2005, with increases in state appropriations and child enrollment levels each year. Among the 13 states designated as being in the West region by the U.S. Census, only seven

currently offer a state prekindergarten program (Barnett et al., 2008). Within this region, New Mexico is now second only to Colorado (13% vs. 16%) in terms of the percentage of 4-year-olds in each state who are enrolled. Table 1 provides more information about annual funding and enrollment levels for the New Mexico PreK initiative.

Table 1. Funding and Enrollment Levels for New Mexico PreK Since 2005

School Year	State Appropriation	Child Enrollment (Number of 4-Year-Olds)
2005-2006	$4,950,000	1,538
2006-2007	$7,990,000	2,195
2007-2008	$14,000,000	3,570
2008-2009	$19,090,000	4,623
2009-2010 (anticipated)	$19,842,400	4,930

Like the New Mexico PreK initiative itself, the NIEER evaluation of New Mexico PreK began during the 2005-2006 school year. This is the third report examining the impacts of New Mexico PreK on young children's language, literacy, and mathematics skills as they enter kindergarten. Two previous reports using a similar methodology have examined the effects of New Mexico PreK programs on children entering kindergarten in fall 2006 (Hustedt, Barnett, & Jung, 2007) and fall 2007 (Hustedt, Barnett, Jung, & Figueras, 2008). Each of these reports presents statistically significant and meaningful impacts of the PreK initiative on children's language, literacy, and math skills. The current report focuses on the effects of New Mexico PreK programs on children who entered kindergarten in fall 2008, and also presents findings using a multi-year data set that combines results from fall 2006, fall 2007, and fall 2008.

Methods

The Regression-Discontinuity Approach

Our research model is based upon a regression-discontinuity design (RDD) in order to reduce selection bias. In state prekindergarten evaluations, the effects of an initiative are often estimated by comparing test scores of children who attended the pre-K program with the scores of similar children who did not. However, it becomes increasingly difficult to find a comparable group of children who did not attend, as statewide programs become more widely available. Even where programs target only a subset of children (such as those from low-income families), the children who attend preschool are different from children who do not. Preschool programs

contribute to these differences by targeting different groups of children, but differences also come about because only some eligible families choose to enroll their children. In sum, selection bias can be a problem because programs select children and families select programs, ultimately leading to differences between pre-K participants and non-participants.

In the New Mexico study, our solution to the problem of selection bias involves comparing two groups of children, where both groups of children have enrolled in New Mexico PreK. The RDD comparisons rely upon New Mexico's stringent age cut-off for enrollment eligibility (August 31) to define the two groups. This concept is easier to understand by providing an extreme example: consider two children who differ only in that one was born the day before the age cut-off and the other the day after. When both children are about to turn 5 years old, the slightly younger child will enter PreK. The slightly older child will enter kindergarten having already completed PreK. If both children are tested at the beginning of the school year, the difference in their scores provides an unbiased estimate of effect of PreK. If only children with birthdays one day on either side of the age cut-off were included in a study, the sample size would be unreasonably small. However, the RDD approach can be applied to wider age ranges around the cut-off. In fact, all children entering kindergarten having completed New Mexico PreK, and all children beginning New Mexico PreK the same year, can be included in the analyses. The RDD approach has been used recently in a growing body of research examining the effects of state-funded pre-K programs in Oklahoma (Gormley et al., 2005), New Jersey (Frede, Jung, Barnett, Lamy, & Figueras, 2007), Arkansas (Hustedt, Barnett, Jung, & Thomas, 2007), and several other states (Wong, Cook, Barnett, & Jung, 2008).

As mentioned previously, the RDD approach has been used in the New Mexico PreK evaluation each year since fall 2006, and this report presents results from a third set of RDD analyses. Each set of RDD analyses investigates the impacts of participating in the New Mexico PreK initiative at age 4 on children's academic skills at kindergarten entry. Appendix A provides more details about the how the RDD approach is used in this study.

Sampling Strategy

In fall 2008, we began our sample selection process by first identifying all sites that had New Mexico PreK programs by the beginning of the 2007-2008 school year and that were continuing operations into the 2008-2009 school year. (In communities offering PreK for the first time in fall 2008, we did not expect to locate sufficient numbers of kindergartners who had completed PreK the previous school year.) At each identified PreK site we randomly selected a pre-specified number of children to participate in our research, based on the proportion of New Mexico PreK children statewide who were enrolled at that particular site. PreK children were then selected from class enrollment lists using a procedure to ensure randomness, and were assessed at their PreK site until we had completed the designated number of assessments.

We also chose a corresponding number of kindergartners for each of the selected fall 2008 PreK sites. We identified kindergartners who had participated in New Mexico PreK the previous year using the state's PreK enrollment list from 2007-2008, and randomly selected children from that list. Children were tracked to their current elementary schools using information about their anticipated kindergarten destinations collected at the end of the previous

school year by PED and CYFD, and compiled by the state Office of Education Accountability. Current kindergarten students were then assessed at their elementary schools.

New Mexico-based child assessors – trained by NIEER and working under the supervision of researchers from New Mexico State University – visited each sampled PreK site as well as kindergarten sites where former PreK participants had enrolled. Research staff conducted child assessments as early as possible in the school year.

The Fall 2008 Sample

Our RDD methodology compares the two groups of children mentioned above. The group of kindergartners who attended New Mexico PreK the previous year (2007-2008) is considered the *Preschool* group, or the experimental group. Children who received some form of early care or education instead of New Mexico PreK at age 4 were not included in this group. The second group of children is considered the *No Preschool* group, or the control group. This group is referred to as the No Preschool group even though these children enrolled in PreK during 2008-2009, because at the time of the assessments they were starting the preschool year and had not experienced the preschool "treatment" yet. By comparing these two groups based on data collected in fall 2008, we are able to estimate the effects of attending PreK during the 2007-2008 school year.

In fall 2008, the No Preschool group included 680 children from New Mexico PreK classrooms across the state. The Preschool group included 653 children from kindergarten classrooms across the state. The total New Mexico sample size was 1,333 children.

The total fall 2008 sample was 53.9% female. Children's home languages were: English only, or English plus another language, 84.2%; Spanish only, 14.8%; and other languages, 0.3%. The percentage of children in each ethnic category was: Hispanic, 57.6%; White, 23.9%; Native American, 14.5%; Black, 1.7%; Asian, 0.4%; and Other, 0.7%.

Ethnicities of participants in our study generally reflect those of the population of children participating in the New Mexico PreK program. During the 2008-2009 school year, the percentage of all New Mexico PreK children in each ethnic category was: Hispanic, 61.8%; Caucasian, 21.1%; American Indian and Alaska Native, 13.9%; Black, 1.9%; and Asian, 1.3%.

For purposes of comparison, New Mexico-specific estimates from the U.S. Census (Bureau of Business and Economic Research, 2009) show that the percentages of New Mexico children ages birth to 5 in each ethnic category were as follows for 2008: Hispanic, all races, 56.3%; White, 27.1%; American Indian and Alaska Native, 10.5%; two or more races, non-Hispanic, 2.4%; Black, 2.4%; and Asian, 1.2%.

Research Instruments

Child outcome measures in the New Mexico PreK study focus on vocabulary, math, and early literacy skills. All measures selected for this study were chosen so that child assessments could be conducted in either English or Spanish. Specific details about the measures and our

protocols for determining the language of assessment are provided below. Also, recognizing that some children in New Mexico may need assistance with the cultural context of the standardized assessment instruments, we developed an additional protocol so that a "cultural broker" could be present for children who might have difficulties with the instruments due to cultural differences. Teachers were asked to identify children who needed this type of assistance and to identify a cultural broker who could be present during testing. We requested that the cultural broker be someone conversant with the child's own culture, and ideally a school employee familiar to the child. The child assessment battery used during the 2008-2009 school year was unchanged from the assessment battery used during the previous school year.

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 predictive of general cognitive abilities and is a direct measure of vocabulary size. The rank order of item difficulties is highly correlated with the frequency with which words are used in spoken and written language. This test is adaptive (to avoid floor and ceiling problems), 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 or test-retest reliabilities. The TVIP is appropriate for measuring growth in Spanish vocabulary for bilingual students and for monolingual Spanish speakers. The results of these tests are found to be strongly correlated with school success. Raw scores are reported here.

All children in this study were initially administered the PPVT, regardless of their home language, to get a sense of their receptive vocabulary skills in English. Children who spoke some Spanish were also subsequently administered the TVIP. The testing session was then continued, with all additional measures administered in either English or Spanish, depending upon which language the child's teacher designated as his or her best testing language.

Math Skills. Children's early math skills were measured using the Woodcock-Johnson Tests of Achievement, 3rd Edition (WJ-III; Woodcock, McGrew & Mather, 2001) Applied Problems subtest. For children whose best testing language was Spanish, 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 here.

Early Literacy. Print knowledge was measured using the Print Knowledge subtest of the Test of Preschool Early Literacy (TOPEL; Lonigan, Wagner, Torgesen, & Rashotte, 2007). The TOPEL is the published, normed version of the unpublished Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP; Lonigan, Wagner, Torgesen & Rashotte, 2002), which was used in the New Mexico PreK evaluation during the 2006-2007 school year. As the TOPEL has not yet been published in Spanish, we continued using the Spanish version of the Pre-CTOPPP with children whose best testing language was Spanish.

The TOPEL has been used with both middle- and low-income samples, and subtests are reported to have good to excellent reliability. Print Knowledge items measure whether children recognize individual letters and letter-sound correspondences, and whether they differentiate

words in print from pictures and other symbols. Percentages of items answered correctly out of the total 36 Print Knowledge subtest items are reported here.

Findings

A statistical description of the fall 2008 sample by group (entering preschool and entering kindergarten) is provided in Tables 2 and 3. As shown in Table 2, the groups are similar in their demographic characteristics, indicating that our sampling process was successful in obtaining comparable experimental and control groups. As shown in Table 3, the test scores are different between the two groups, as would be expected, though care must be taken not to interpret the simple differences between the groups' test scores as an estimate of the PreK program's impact.

Table 2. Statistical Description of the Sample by Group, for Children Entering PreK and Children Entering Kindergarten from PreK in Fall 2008

	Entering PreK	Kindergarten
Number in group	680	653
Girls (%)	55.1	52.5
Ethnicity (%)		
White	25.1	22.7
Hispanic	56.6	58.7
Native American	14.4	14.5
Other/missing	3.8	4.1
Home Language		
English, or English + another language	84.1	84.4
Spanish only	14.7	14.9
Other	0.4	0.2
Assessment conducted only in English (%)	86.8	89.7
Age (in months) when assessed (Mean/SD)	55.20	67.39
	3.82	4.02

Note: SD = Standard Deviation

Table 3. Assessment Scores by Group, for Children Entering PreK and Children Entering Kindergarten from PreK in Fall 2008

| Measure | Entering | | | | |
| | PreK | | Kindergarten | | |
	Mean	*SD*	Mean	*SD*	Sample Size
PPVT raw scores	44.78	20.53	61.86	21.06	1,282
WJ-III raw scores	9.42	4.35	14.85	4.39	1,298
TOPEL % correct	30.21	24.10	70.77	26.13	1,298

Note: *SD* = Standard Deviation

Analyses for this study were conducted in STATA (StataCorp, 2005) using raw scores. All standard errors were clustered by classroom. Covariates were included in our statistical models to control for time of assessment, gender, ethnicity, and whether assessments were administered in English or Spanish. In our RDD analyses, there is no *a priori* expectation that the estimated relationship between PreK participation and child outcomes should be linear. Therefore, we estimated linear as well as higher order polynomial forms of the regression equation, as a check against mis-specifying the functional form of the regression line. We conducted squared and cubic transformations of the selection variable (the difference between birth date and cut-off date) and its interaction with the cut-off variable.

Linear, quadratic, and cubic estimates for each of our child outcome measures using fall 2008 data are shown in Table 4. We found that linear models provided the best estimates of relationships between participating in the New Mexico PreK initiative and children's scores on each the three measures. For the remainder of this report, we will focus on the linear estimates when reporting fall 2008 data for the PPVT-III, the WJ-III, and the TOPEL.

Also, in presenting RDD findings we generally emphasize the results using one year as the margin around the kindergarten cut-off date, because this allows us to include the largest sample size. However, we also conduct linear regressions restricting the sample to children born within 3- and 6-month spans before and after the cut-off date. Restricting the sample to observations closest to the cut-point should reduce any potential bias, though the smaller sample sizes increase the standard errors. As shown in Table 4, linear estimates using 3-, 6-, and 12-month margins are similar for this sample.

Table 4. Estimated Effects Based on Functional Form and Margin Around Kindergarten Cut-Off Date in Fall 2008

	Parametric models used in analysis				
	Linear, 12 months	Quadratic, 12 months	Cubic, 12 months	Linear truncated at 3 months	Linear truncated at 6 months
Receptive Vocabulary	**3.42$^+$** **(1.91)**	2.95 (2.54)	2.33 (3.28)	2.99 (3.56)	3.64 (2.49)
Math	**1.86*** **(0.46)**	1.79* (0.66)	1.70* (0.85)	2.06* (0.86)	1.60* (0.64)
Early Literacy	**27.61*** **(2.92)**	29.06* (4.22)	28.56* (5.47)	26.06* (5.74)	29.55* (4.07)

Note: Receptive vocabulary data represent PPVT raw score point increases. Math data represent WJ-III Applied Problems subtest raw score point increases. Early literacy data represent increases in percentage correct on the TOPEL Print Knowledge subtest.
Robust standard errors are shown in parentheses. Boldfaced terms show the selected estimates.
$^+ p < .10.$ $^* p < .05.$

Our primary analyses were "sharp" regression-discontinuity models that included 1,299 children, dropping 34 children (2.6% of the total sample) whose birth-date information appears to be inconsistent with the birth-date cut-off requirements for their PreK or kindergarten programs. When less than 5% of the sample is dropped in this way, exclusion of such cases is thought to have little effect on the result (Judd & Kenny, 1981; Shadish, Cook, & Campbell, 2002; Trochim, 1984). As an additional check to confirm that the findings were similar, we also conducted instrumental variable analyses with all 1,333 children. Results of these analyses were indeed similar, so we will focus on the findings from the "sharp" regression-discontinuity analyses in the remainder of this report.

Child Outcomes from the Third Year of New Mexico PreK

In this section we provide an overview of the impacts of participating in the New Mexico PreK initiative during the 2007-2008 school year. Effects on children's receptive vocabulary, mathematics, and early literacy skills are summarized below, and are shown in graphical form in Appendix A.

The estimated effect of state-funded preschool on children's receptive vocabulary as measured by the PPVT approaches statistical significance ($p < .10$). Attending New Mexico PreK during the 2007-2008 school year was estimated to increase PPVT scores by about 3.42 raw score points at kindergarten entry. This represents an improvement of about 17% of the standard deviation for the control (No Preschool) group.

The estimated effect of New Mexico PreK on children's early math skills as measured by Woodcock-Johnson-III Applied Problems subtest scores was statistically significant for the 2007-2008 school year ($p < .01$). The increase in scores for New Mexico PreK children due to the program is about 1.86 raw score points. This represents an improvement of about 43% of the standard deviation for the control (No Preschool) group.

Finally, the effect of the New Mexico PreK initiative on children's Print Knowledge scores was statistically significant for the 2007-2008 school year ($p < .01$). The effect of the New Mexico PreK initiative on children's gains in Print Knowledge scores is about 28% more items answered correctly. This increase represents approximately 115% of the control (No Preschool) group standard deviation on the Print Knowledge subtest.

<u>Child Outcomes Across the First Three Years of New Mexico PreK</u>

Now that we have gathered RDD data on three separate occasions – representing the impacts of attending New Mexico PreK during each of its first three years of operation – we also have new opportunities to conduct analyses using a pooled, multi-year data set and to examine preliminary trends over time. Table 5 presents linear, quadratic, and cubic estimates for each of our child outcome measures using a data set that includes the results of 3,020 child assessments conducted each fall in 2006, 2007, and 2008. These results are analogous to the findings presented in Table 4, but represent three years of data rather than a single year of data.

Table 5. Estimated Effects Based on Functional Form and Margin Around Kindergarten Cut-Off Date, Using Multi-Year Dataset

	Parametric models used in analysis				
	Linear, 12 months	Quadratic, 12 months	Cubic, 12 months	Linear truncated at 3 months	Linear truncated at 6 months
Receptive Vocabulary	4.39* (1.29)	4.74* (1.82)	**5.44*** **(2.38)**	6.56* (2.48)	5.82* (1.72)
Math	**1.63*** **(0.32)**	1.58* (0.47)	1.49* (0.63)	2.04* (0.63)	1.69* (0.45)
Early Literacy	**24.43*** **(1.96)**	22.93* (2.80)	21.82* (3.71)	22.88* (3.77)	24.03* (2.75)

Note: Receptive vocabulary data represent PPVT raw score point increases. Math data represent WJ-III Applied Problems subtest raw score point increases. Early literacy data represent increases in percentage correct on the TOPEL Print Knowledge subtest.
Robust standard errors are shown in parentheses. Boldfaced terms show the selected estimates.
*$p < .05$.

Using this pooled data set across all three years, the estimated effect of New Mexico PreK on children's receptive vocabulary was statistically significant ($p < .05$). Attending New Mexico PreK was estimated to increase PPVT scores by about 5.44 raw score points at kindergarten entry. This represents an improvement of about 25% of the standard deviation for the control (No Preschool) group.

The estimated effect of New Mexico PreK on children's early math skills as measured by the Woodcock-Johnson-III Applied Problems subtest scores was also statistically significant ($p < .01$). The increase in scores for New Mexico PreK children due to the program is about 1.63 raw score points, representing an improvement of about 37% of the standard deviation for the control (No Preschool) group.

Table 6. Estimated Effects Based on Functional Form and Margin Around Kindergarten Cut-Off Date, for Each Year of New Mexico PreK

	Parametric models used in analysis				
	Linear, 12 months	Quadratic, 12 months	Cubic, 12 months	Linear truncated at 3 months	Linear truncated at 6 months
Receptive Vocabulary					
Year 1	**7.82***	7.69*	6.03	9.22	7.72*
Year 2	**5.55***	8.26*	13.19*	12.08*	10.26*
Year 3	**3.42+**	2.95	2.33	2.99	3.64
Multi-year	4.39*	4.74*	**5.44***	6.56*	5.82*
Mathematics					
Year 1	**1.64***	1.29	1.59	1.79	1.80+
Year 2	1.88*	1.99*	**2.26***	2.61*	2.34*
Year 3	**1.86***	1.79*	1.70*	2.06*	1.60*
Multi-year	**1.63***	1.58*	1.49*	2.04*	1.69*
Early Literacy					
Year 1	**26.08***	26.75*	23.71*	29.13*	26.24*
Year 2	21.36*	**13.66***	13.36*	15.15*	17.39*
Year 3	**27.61***	29.06*	28.56*	26.06*	29.55*
Multi-year	**24.43***	22.93*	21.82*	22.88*	24.03*

Note: Receptive vocabulary data represent PPVT raw score point increases. Math data represent WJ-III Applied Problems subtest raw score point increases. Early literacy data represent increases in percentage correct on the Pre-CTOPPP Print Awareness subtest (Year 1) or TOPEL Print Knowledge subtest (Years 2 and 3).
Year 1 shows the impact of attending New Mexico PreK during the 2005-06 school year, Year 2 shows the impact of attending New Mexico PreK during the 2006-2007 school year, and Year 3 shows the impact of attending New Mexico PreK during the 2007-2008 school year.
Boldfaced terms show the selected estimates.
$^+p < .10$. $^*p < .05$.

The estimated effect of the New Mexico PreK initiative on children's Print Knowledge scores was statistically significant as well ($p < .01$). The effect of the New Mexico PreK initiative on children's gains in Print Knowledge scores is about 24% more items answered correctly. This increase represents approximately 102% of the control (No Preschool) group standard deviation on the Print Knowledge subtest.

While the findings just described use one year as the margin around the kindergarten cut-off date, our results are also extremely positive using 3- and 6-month spans before and after the cut-off date. Restricting the sample to observations closest to the cut-point should reduce any potential bias. These estimates of the impacts of New Mexico PreK are likely to be more precise, although the smaller sample sizes also increase the standard errors (see Table 5).

Table 6 presents the effects of the New Mexico PreK initiative on children's receptive vocabulary, math, and early literacy scores separately for each of the first three years of the initiative, in order to facilitate comparisons across different years of the PreK program. The estimated impacts of New Mexico PreK are consistent across each year of the study for mathematics and early literacy. However, these preliminary trends appear to show a decline in the impacts of New Mexico PreK on children's receptive vocabulary scores in the third year of the initiative. As noted previously, the receptive vocabulary findings for the 2007-2008 school year do not reach statistical significance, though when the multi-year data set is used, the findings in this content area are statistically significant.

Receptive Vocabulary, Math, and Early Literacy Results for CYFD and PED

In fall 2008 we increased our sample size to provide additional statistical power needed to separately examine the impacts of participating in PreK sites offered through CYFD and those offered through PED. These analyses were based on the subgroups of 703 children who attended New Mexico PreK at PED sites and 619 children who attended New Mexico PreK at CYFD sites. The results of those analyses, and comparisons with analyses from the overall New Mexico PreK sample are shown in Table 7.

Table 7. Estimated Effects Disaggregated for PED and CYFD Sites and Overall in Fall 2008

Measure	PED	CYFD	All NM PreK
Receptive Vocabulary			
Linear	2.70	5.38^{+}	3.42^{+}
Math			
Linear	1.39^{*}	2.73^{*}	1.86^{*}
Early Literacy			
Linear	28.63^{*}	28.04^{*}	27.61^{*}

Note: Receptive vocabulary data represent PPVT raw score point increases. Math data represent WJ-III Applied Problems subtest raw score point increases. Early literacy data represent increases in percentage correct on the TOPEL Print Knowledge subtest.
$^{+}p < .10.$ $^{*}p < .05.$

Like the New Mexico PreK initiative as a whole, CYFD and PED programs produced statistically significant impacts on measures of children's math and early literacy skills for PreK participants who entered kindergarten in fall 2008. Statistically significant impacts were not found for the receptive vocabulary measure, although findings for CYFD and for the New Mexico PreK initiative as a whole approached statistical significance. The findings presented in Table 7 are the result of linear regressions with a 12-month margin around the kindergarten cut-off date. Analyses using 3- and 6-month margins tend to show larger impacts of PreK, particularly for PED on receptive vocabulary, though confidence intervals are quite large due to the reduced sample sizes.

In previous years of the study our sample sizes for CYFD and PED were smaller, and thus analyses restricted to each year would have limited power to detect a statistical effect. However, as results from each of the three annual RDD analyses (Table 6) are generally similar, we are able to conduct additional analyses for PED and CYFD using our multi-year data set. The results of those analyses, presented in Table 8, show that impacts of PED and CYFD programs are very similar across the three years of our study.

Table 8. Estimated Effects Disaggregated for PED and CYFD Sites and for the Entire New Mexico PreK Sample, Using Multi-Year Dataset

Measure	PED	CYFD	All NM PreK
Receptive Vocabulary			
Linear	4.51*	4.92*	4.39*
Cubic	5.38+	6.27+	**5.44***
Math			
Linear	1.44*	1.91*	**1.63***
Early Literacy			
Linear	25.75*	23.35*	**24.43***

Note: Receptive vocabulary data represent PPVT raw score point increases. Math data represent WJ-III Applied Problems subtest raw score point increases. Early literacy data represent increases in percentage correct on the TOPEL Print Knowledge subtest.
+ $p < .10$. * $p < .05$.

Discussion

During a period of rapid expansion in New Mexico PreK, the state-funded prekindergarten initiative continues to produce positive impacts on children's early learning when they start kindergarten, as illustrated by the findings in this report. The effects of attending New Mexico PreK during its third year of implementation are comparable to results from the PreK initiative's first and second years (Hustedt et al., 2007; Hustedt et al., 2008) in terms of impacts on early literacy and math. The impacts that year on children's language skills are smaller. Additional analyses across the first three years of the New Mexico PreK initiative continue to show statistically significant impacts in the areas of language, literacy, and math.

Improvements in children's early learning skills such as those produced by New Mexico PreK are meaningful, as they can lead to increased success in school.

Figure 1 shows effect sizes for each content area included in our child assessment battery, with estimates for each year of our study as well as an overall estimate encompassing data from all three years. Effect sizes were calculated by comparing the regression coefficient to the standard deviation for the control group (analogous to Glass's delta), and help to standardize estimated effects of the PreK program across different types of measures and across successive years of the study. The effect sizes for our receptive vocabulary measure signal a decrease in average scores for children attending consecutive years of the New Mexico PreK program. Further analyses and more successive years of data will be needed to determine if this potential trend is meaningful. There are no apparent trends over time in the effect sizes for either print knowledge or math. The print knowledge results in Year 2 seem to be an anomaly that may, in part, be due to our selection of a non-linear model to represent print knowledge results for the 2006-2007 school year.

Figure 1. The Effect of the New Mexico PreK Initiative on Children's Scores across Measures
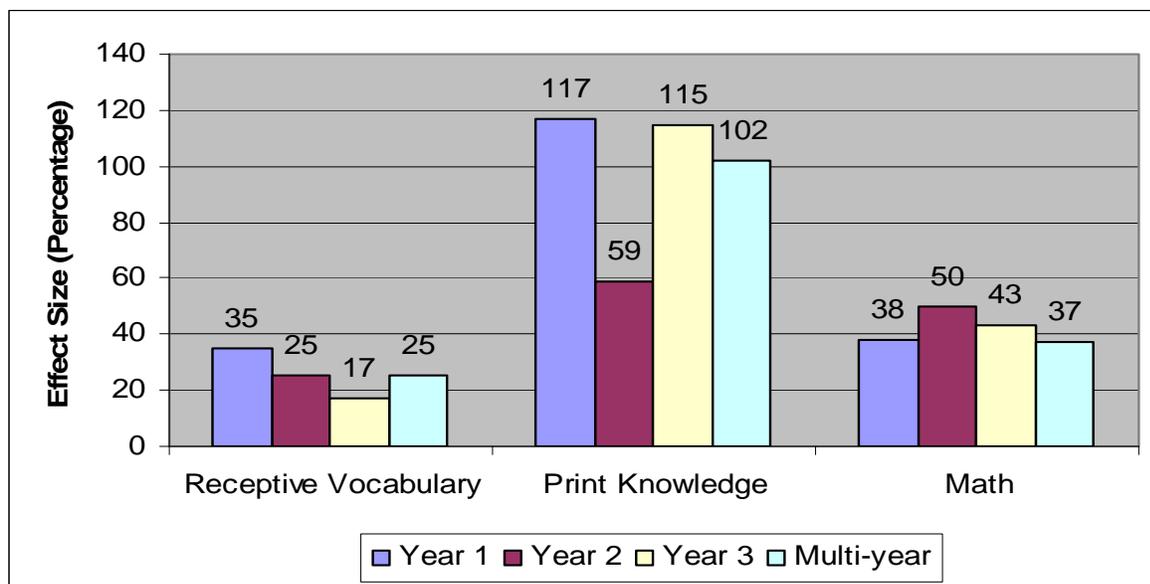
Note: Year 1 shows the impact of attending New Mexico PreK during the 2005-2006 school year, Year 2 shows the impact of attending New Mexico PreK during the 2006-2007 school year, and Year 3 shows the impact of attending New Mexico PreK during the 2007-2008 school year.

Effect sizes for New Mexico PreK are comparable in magnitude to those that have been reported for other well-regarded state preschool initiatives, such as the widely researched initiative in Oklahoma (e.g., Gormley et al., 2005). As emphasized in our previous reports, though, New Mexico's situation is uncommon in that the state began a program evaluation during the first year of the initiative. This provides a unique opportunity to document the progress by New Mexico PreK as it becomes more established and becomes more available across the state. However, caution in interpreting these evaluation findings is also needed.

While the impacts found in this study are precursors to later success in school and may ultimately lead to the types of economic benefits found in long-term studies of preschool education (Schweinhart, Montie, Ziang, Barnett, Belfield, & Nores, 2005; Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002; Reynolds et al., 2002), New Mexico PreK is still a new initiative and should continue to be viewed as a work in progress. The Oklahoma study and the previous state prekindergarten evaluations conducted by NIEER have focused on initiatives that had been well established for a number of years before these studies were started.

The larger sample of children included in our fall 2008 data collection cycle allow us to better understand the impacts of New Mexico PreK programs offered in different types of settings. We increased our sample sizes of children enrolled in both PED and CYFD programs with specific goals of increasing statistical power and more closely examining impacts of PreK programs administered by PED and CYFD. Results from separate sets of analyses conducted for PED and CYFD programs largely mirror the overall results for the New Mexico PreK initiative as a whole (e.g., analyses where we did not distinguish between programs based on the lead administrative agency). Also, pooled findings across all three years of data are consistent with this finding from the 2007-2008 school year, providing more confidence in the conclusion that programs operated by CYFD and PED have similar impacts on children at the current stage of implementing the New Mexico PreK initiative.

Program quality is an important factor in the implementation of prekindergarten initiatives. The NIEER New Mexico PreK evaluation has also examined the observed quality of classrooms by conducting classroom visits. These data have been gathered during the spring of each school year since the New Mexico PreK initiative began. Our classroom observations have used a standard battery of tools focusing on overall classroom quality (the ECERS; Harms, Clifford, & Cryer, 2005), early language and literacy support (the SELA; Smith, Davidson, & Weisenfeld, 2001), and mathematics support (the PCMI; Frede, Weber, Hornbeck, Stevenson Boyd, & Colon, 2005). General findings from these classroom observations suggest that overall quality is good, that classroom support for early language and literacy is average, and that there is considerable room for improvement in classroom support for early mathematics (Hustedt & Barnett, 2008). In each year of the New Mexico PreK evaluation, the mean classroom quality scores using all three observation tools have been similar for programs offered by CYFD and those offered by PED.

While overall quality scores for the New Mexico PreK initiative have been relatively steady over the course of this study, one potential consequence of the initiative's rapid expansion is that classroom quality may begin to decline over time as new programs are added. Reductions in classroom quality could result in PreK having smaller impacts on children who participate that are ultimately reflected in child outcome measures. Further interpretation of our classroom quality data and examination of potential linkages between classroom quality and child outcome data will be provided in the final report for this program evaluation cycle. Possible analyses that may provide further insight into these data include examination of additional classroom factors, such as teacher qualifications. Also, it will be possible to look at classroom quality in our initial sample of classrooms to examine trends in classroom quality over time.

The current report is the final of three regression-discontinuity reports from an initial 2005-2009 program evaluation of New Mexico PreK. During the course of this evaluation, the New Mexico PreK initiative has grown substantially, from an enrollment of about 1,500 children when it began in 2005, to an anticipated enrollment of nearly 5,000 children in 2009. Further growth is anticipated, and in extremely difficult budget times New Mexico was one of 14 states to propose a budget increase for state-funded prekindergarten in fiscal year 2010 (Pre-K Now, 2009). In the context of this ongoing growth, and as additional teacher education requirements begin to be phased in, it is important to continue documenting the impacts of New Mexico PreK on young children. A second program evaluation of New Mexico PreK is planned for 2009-2013, including annual regression-discontinuity reports examining children's progress as they enter kindergarten. The new study will also include a longitudinal component, investigating the impacts of participating in New Mexico PreK as its first cohort of participants reaches third grade. This study design will provide additional insight regarding the short and longer-term effects of the New Mexico PreK initiative as it continues to expand.

References

Barnett, W. S. (1996). *Lives in the balance: Age 27 benefit-cost analysis of the High/Scope Perry Preschool Program*. Ypsilanti, MI: High/Scope Press.

Barnett, W. S. (2002). Early childhood education. In A. Molnar (Ed.), *School reform proposals: The research evidence* (pp. 1-26). Greenwich, CT: Information Age Publishing, Inc.

Barnett, W. S. (2003). Better teachers, better preschools: Student achievement linked to teacher qualifications. *Preschool Policy Matters, 2*. New Brunswick, NJ: National Institute for Early Education Research.

Barnett, W. S., Epstein, D. J., Friedman, A. H., Stevenson Boyd, J., & Hustedt, J. T. (2008). *The state of preschool 2008: State preschool yearbook*. New Brunswick, NJ: National Institute for Early Education Research.

Barnett, W. S., Friedman, A. H., Hustedt, J. T., & Stevenson Boyd, J. (2009). An overview of prekindergarten policy in the United States: Program governance, eligibility, standards, and finance. In R. C. Pianta & C. Howes (Eds.), *The Promise of Pre-K* (pp. 3-30). Baltimore, MD: Brookes Publishing.

Barnett, W. S., Hustedt, J. T., Hawkinson, L. E., & Robin, K. B. (2006). *The state of preschool 2006: State preschool yearbook*. New Brunswick, NJ: National Institute for Early Education Research.

Barnett, W. S., & Yarosz, D. J. (2007). Who goes to preschool and why does it matter? *Preschool Policy Matters, 15*. New Brunswick, NJ: National Institute for Early Education Research.

Bureau of Business and Economic Research, University of New Mexico. (2009). *State and county population estimates by age, sex, race and Hispanic origin from the Census Bureau*. Retrieved July 21, 2009, from http://www.unm.edu/~bber/demo/coestchar.htm

Campbell, F. A., Ramey, C. T., Pungello, E. P., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian Project. *Applied Developmental Science, 6,* 42-57.

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.

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.

Gilliam, W., & Zigler, E. (2004). *State efforts to evaluate the effects of prekindergarten: 1977 to 2003*. Unpublished manuscript: Yale University Child Study Center.

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.

Gormley, W.T., Jr., Phillips, D., & Gayer, T. (2008). Preschool programs can boost school readiness. *Science, 320,* 1723-1724.

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. (2008, November). *The New Mexico PreK initiative: Impacts on children's school readiness and recent classroom quality results*. Presentation to the Legislative Education Study Committee, State of New Mexico, Santa Fe, NM.

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.

Hustedt, J. T., Barnett, W. S., Jung, K., & Figueras, A. (2008). *Impacts of New Mexico PreK on children's school readiness at kindergarten entry: Results from the second year of a growing initiative*. New Brunswick, NJ: National Institute for Early Education Research.

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. Available at http://nieer.org/resources/research/ArkansasYear1.pdf

Judd, C. M., & Kenny, D. A. (1981). *Estimating the effects of social interventions*. New York: Cambridge University Press.

Lonigan, C. J., Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (2007). *Test of Preschool Early Literacy (TOPEL)*. Austin, TX: Pro-Ed.

Lonigan, C. J., Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (2002). *Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP)*. Tallahassee, FL: Department of Psychology, Florida State University.

Masse, L. N., & Barnett, W. S. (2002). *A benefit-cost analysis of the Abecedarian Early Childhood Intervention*. New Brunswick, NJ: National Institute for Early Education Research.

Mitchell, A.W. (2001). *Education for all young children: The role of states and the federal government in promoting prekindergarten and kindergarten*. FCD Working Paper. New York: Foundation for Child Development.

Pre-K Now. (2009). *Leadership matters: Governors' pre-K proposals fiscal year 2010*. Washington, DC: Author.

Pre-Kindergarten Act, NMSA 1978 § 32A-23 (2005).

Reynolds, A. J., Temple, J.A., Robertson, D.L., & Mann, E.A. (2002). *Age 21 cost-benefit analysis of the Title I Chicago Child-Parent Centers.* (Discussion Paper no. 1245-02). Madison, WI: Institute for Research on Poverty. Available at http://www.irp.wisc.edu/publications/dps/pdfs/dp124502.pdf.

Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). *Lifetime effects: The High/Scope Perry Preschool study through age 40* (Monographs of the High/Scope Educational Research Foundation, 14). Ypsilanti, MI: High/Scope Educational Research Foundation.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin Company.

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 Publications.

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 A: Applying the RDD Approach to the New Mexico PreK Study

The regression-discontinuity design (RDD) takes advantage of a strict kindergarten 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 comparing children who attended and did not attend the New Mexico PreK initiative (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 New Mexico PreK. One group has completed PreK 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 cut-off. The RDD creates groups that *at the margin* differ only in that some were born a few days before the age cut-off and others a few days after the cut-off. When these children are about to turn 5 years old, the slightly younger children will enter PreK and the slightly older children will enter kindergarten having already attended PreK. If all of the children are tested at that time, the difference in their scores can provide an unbiased estimate of the effect of the preschool initiative under reasonable circumstances. However, if the sample was restricted to children with birthdays near the age eligibility cut-off, the total sample size would be too 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-cut-off sample is used to model the relationship prior to treatment. The post-cut-off sample is used to model the relationship after the treatment. This approach can be applied to wider age ranges around the cut-off, though 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 a variety of models (different functional forms for the equation) to see which model best fits the data, 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 cut-off.

To identify the proper functional form for our RDD analyses, we conducted a graphical analysis and a series of parametric regressions using alternate specifications. We begin with graphs for each child outcome measure, shown in Figures A1-A3. Two types of lines are fitted onto scatterplots on each side of the cut-offs. The first plot in each figure depicts a linear regression line, and the second shows a non-parametric regression line based on locally weighted smoothing, called Lowess. This strategy can be useful 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, we obtain a smoothed value through weighted regressions involving only observations within a local interval, with observations closer to y_i weighted most heavily.

Each plot in Figures A1- A3 shows an estimated regression line for children's predicted test scores by age, measured by the number of days their birth date is from the program enrollment cut-off date. The section of the line to the left of the enrollment cut-off date (shown as "0" in the figures) represents scores of children beginning the state pre-K program, while the section of line to the right of the enrollment cut-off date represents scores for children entering

kindergarten. The discontinuity in the regression line at the cut-off date shows the estimated effect of New Mexico PreK.

Next, we run a series of regressions to obtain parametric estimates of the treatment effect. In order to describe the effect of PreK participation on child outcomes, we model children's vocabulary, math, and early literacy scores. For the ith child in classroom j, the relevant equation is:

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

In this equation, Y_{ij} is child i's outcome, X_{ij} is a vector of child characteristics, Pre-K_{ij} is a dichotomous indicator variable such that $T=1$ for the PreK "treatment" and $T=0$ for no treatment, and $g(AV)_{ij}$ is a smooth function of the continuous assignment variable. We check the robustness of our estimates by considering alternative specifications for $g(AV)_{ij}$, including polynomials and interaction terms. We determine the order of the polynomial approximation to the $g(AV)_{ij}$ function 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. As a final parametric check on functional form, we truncate the dataset to include only observations near the cut-off. In placing greater weight on these observations, we eliminate the influence of extreme assignment variable values that often play a disproportionate role in mis-specifying functional form. We rerun the parametric analyses including only children who have birthdates within 6 and then 3 months on either side of the enrollment cut-off.

Graphical and parametric analyses provide evidence that the response function was linear for receptive vocabulary, quadratic for early literacy outcomes, and cubic for mathematics outcomes. Estimates were robust with respect to narrowing the time window around the birthdate cut-off to 6 and 3 months.

**Figure A1. Linear and Lowess Plots of PPVT Receptive Vocabulary Results**
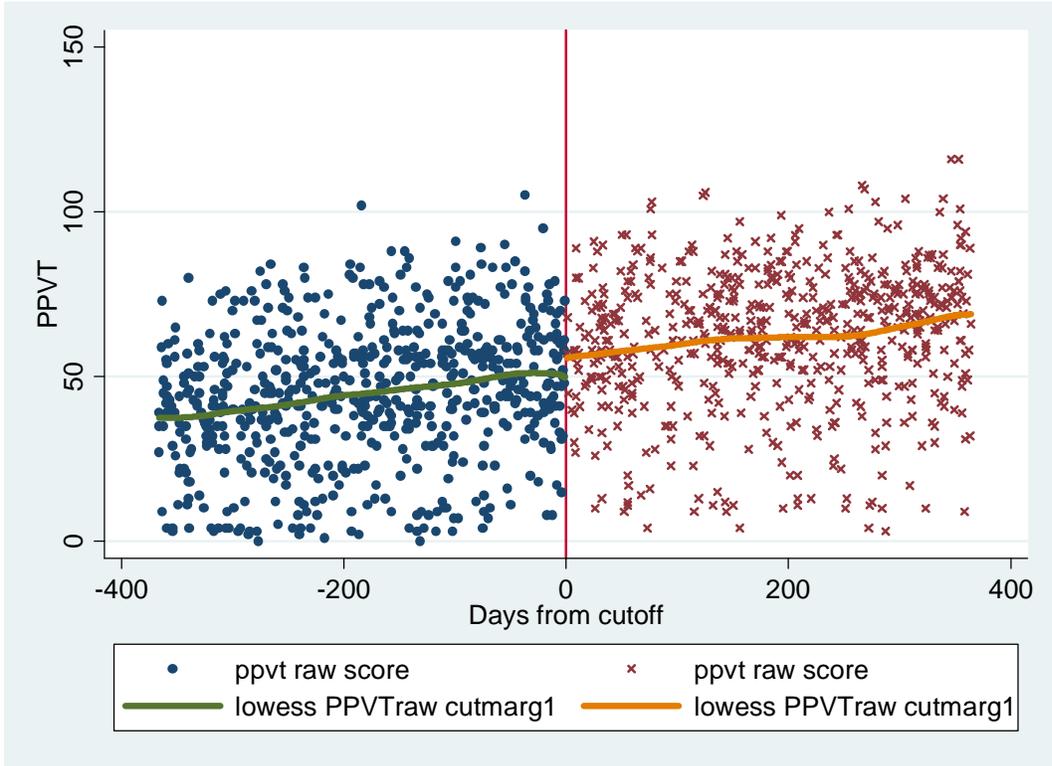
(1) Linear Plot
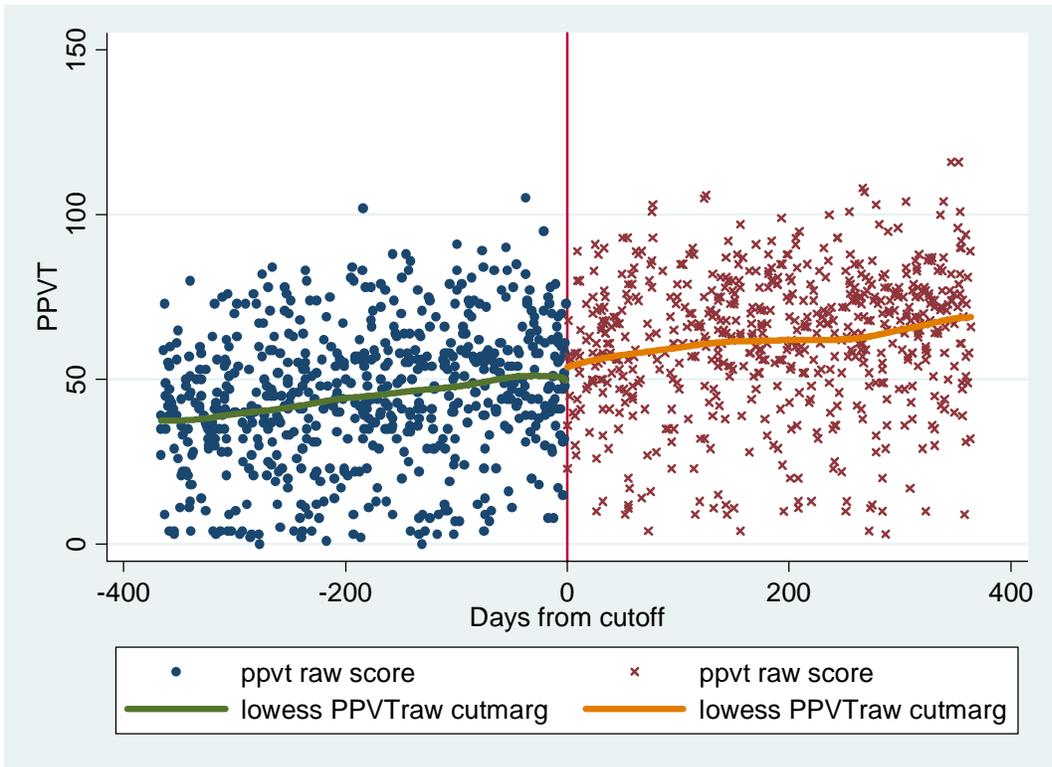
(2) Lowess Plot

Figure A2. Linear and Lowess Plots of WJ-III Applied Problems Results
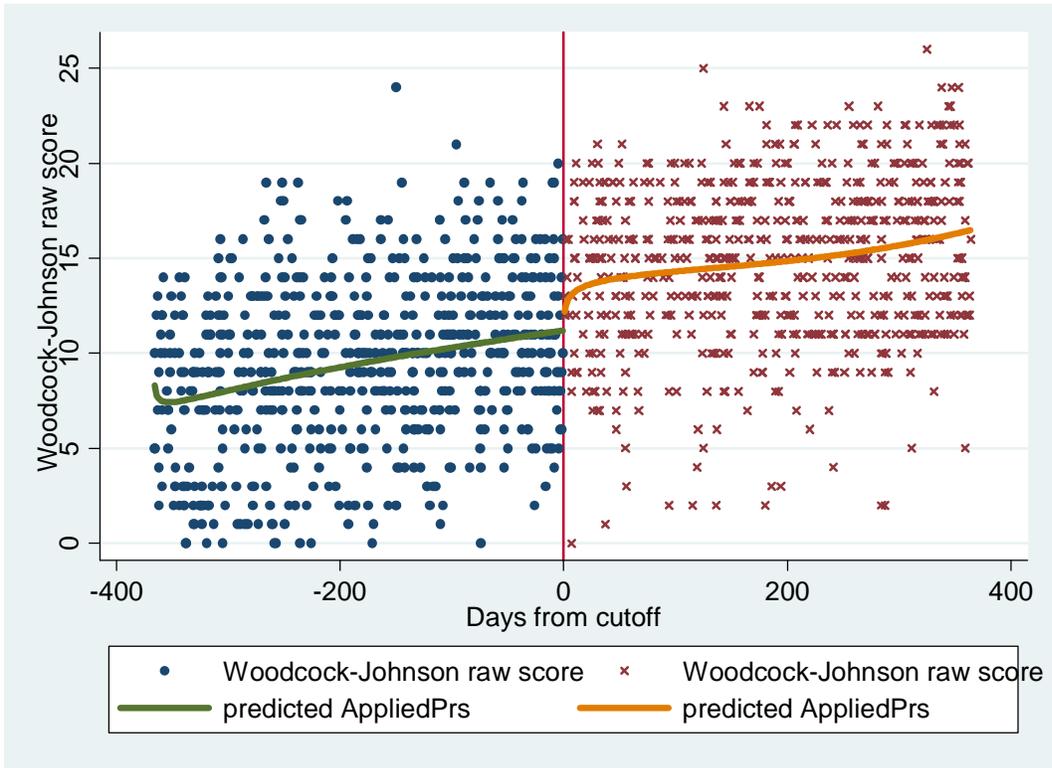
(1) Linear Plot
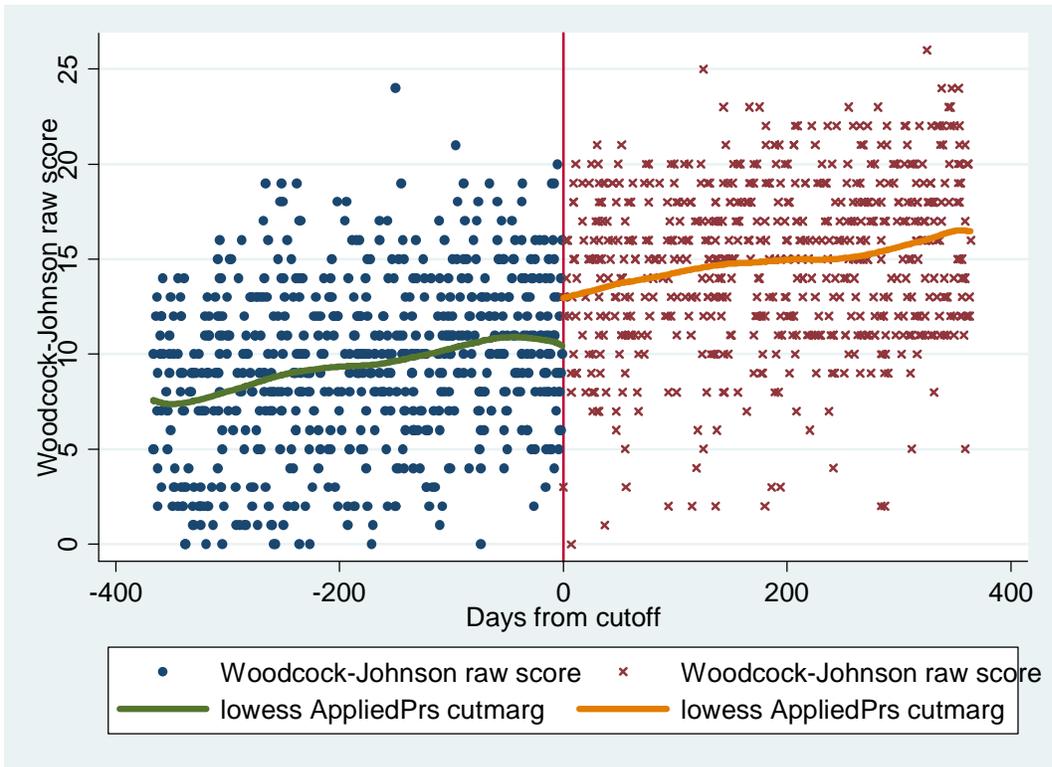
(2) Lowess Plot

Figure A3. Linear and Lowess Plots for TOPEL Print Knowledge Results

(1) Linear Plot

(2) Lowess Plot