

**The Effects of State Prekindergarten Programs on Young
Children's School Readiness in Five States**

W. Steven Barnett
Cynthia Lamy
Kwanghee Jung

The National Institute for Early Education Research
Rutgers University

December, 2005

Funding for this study was provided by The Pew Charitable Trusts, the New Jersey Department of Education as part of the Early Learning Improvement Consortium, the Schumann Fund for New Jersey, the Geraldine R. Dodge Foundation, the Joyce Foundation of Chicago, Illinois, the South Carolina Department of Education and First Steps to School Readiness Program of South Carolina.

Many thanks are due to our colleagues at NIEER who helped support this study, especially Pat Ainsworth, Chris Gilbert, Louise Hornecker, Rena Robin, Carol Shipp, Jessica Thomas, Colleen Vandervort, Valerie Werstler, and Rose Yang.

We could not have accomplished this work without the diligent efforts of our research partners:

The High/Scope Educational Research Foundation

Dr. Larry Schweinhart, Dr. Marijata Daniel-Echols, Carol Markley, and their field staff;

The Early Childhood Collaborative at Oklahoma University

Dr. Loraine Dunn, Dr. Stacy Dykstra, Tamra Crabtree, and their field staff;

South Carolina University

Dr. Nancy Freeman, Sarah Crawford, and their field staff;

Marshall University

Dr. Mary Jo Graham, Paula Strawder, Elizabeth Coffey, and their field staff.

We would like to thank the Michigan, New Jersey, Oklahoma, South Carolina and West Virginia offices of Early Childhood Education, especially Dr. Lindy Buch, Dr. Ellen Frede and Dr. Ellen Wolock, Dr. Ramona Paul, Dr. Linda Mims and Dr. Cathy Jones, for their vision and support.

We are most appreciative of the many school administrators and teachers who supported this work. We are indebted to several fine reviewers who made helpful comments on previous versions of this paper. We are especially grateful for our sample preschool and kindergarten children, who took time out of their busy days to help enlighten us.

The Effects of State Prekindergarten Programs on Young Children's School Readiness in Five States

Executive Summary

This study estimated the effects of five state-funded preschool programs on entering kindergartners academic skills using a rigorous research design. Receptive vocabulary, early literacy and math skills were assessed in a sample of 5071 children from Michigan, New Jersey, Oklahoma, South Carolina, and West Virginia. We found these state-funded preschool programs to have statistically significant and meaningful impacts on children's early language, literacy and mathematical development, with some evidence of an enhanced program effect for print awareness skills for children in low-income families.

Specific findings are as follows:

1. State-funded preschool produces an increase in children's vocabulary scores of nearly 4 raw score points, which equals 31% more growth over the year and an 8 percent increase in children's average vocabulary scores. This improvement translates into an additional four months of progress in vocabulary growth due to the preschool program. This outcome is particularly important because the measure is strongly predictive of general cognitive abilities.
2. Children who attended state-funded preschool scored higher on a test of early math skills. State-funded preschool increased children's math scores by almost 1 and a half raw score points, 44% more growth in a year due to the program and a 13 percent increase in children's average math scores. Skills tested include basic number concepts, simple addition and subtraction, telling time and counting money.
3. State-funded preschool had strong effects on children's understanding of print concepts. The program increased all children's print awareness scores by nearly 17 percentage points, which is 85% more growth over the year and a 39% increase in children's print awareness scores. Children who attended a state-funded preschool program before entering kindergarten know more letters, more letter-sound associations and are more familiar with words and book concepts.
4. We found no significant effects on children's phonological awareness. A relatively new measure was used, and it is difficult to determine whether this result is due to a true lack of program effects. Children in this study appeared to perform well on this test, with or without the preschool program.

Using a sophisticated research design (a regression discontinuity approach) we estimated the gains from one year of state-funded prekindergarten at age 4. Although there were some variations in estimated effects across states (potentially due to differences in populations and availability of other preschool education options), broadly similar effects were found for each of the five state programs. A common element across these programs was that all or nearly all teachers have a four-year college degree with an early childhood specialization.

Introduction

State-funded prekindergarten programs have become increasingly common across the country, having been established to some extent in up to 40 states (Barnett, Robin, Hustedt, & Shulman, 2004). The primary goal of these state-funded programs is to improve the learning and development of young children and improve their preparation for the increasingly rigorous challenges of kindergarten. Effective preschool programs lay a foundation for children's subsequent school success, by imparting the basics – colors, shapes, numbers, letters, how to look at a book, how to get along with classmates, how to live by the rules in school - sending children to kindergarten with solid successes in preschool and the built-in confidence that early success creates.

Earlier studies have shown the value of high-quality, well-funded preschool programs for improving children's short- and long-term success in school and in life (Barnett, 2002). Current state-funded prekindergarten programs are not as well funded as the most effective models studied, but are larger in capacity and serve a more diverse population. The standards and quality of state prekindergarten programs vary greatly. Most state prekindergarten programs target children who are at elevated risk of school failure (often due to poverty), and programs for these children have been the most studied. A few states have recently sought to make prekindergarten education available to all 4-year-olds. Less research has been conducted on the impacts of programs for children who are not economically disadvantaged. This study's contribution to our knowledge about such programs is particularly important.

As the number of state funded prekindergarten programs grows, it is important to determine how effective they are in improving children's learning and development as they enter school. However, it has proven difficult to conduct rigorous evaluations of state preschool programs that provide accurate estimates of effects. The most difficult problem faced by evaluators is possible selection bias due to unmeasured differences between the children who attend state-funded preschool programs and those who do not. In the case of universal programs, it is especially difficult to obtain an adequate comparison group since one is more likely to suspect that some unknown differences in the children and families lead to nonparticipation when a program is freely available to one and all.

Statistical models can be used to estimate program effects, adjusting for known and measured differences between children (and their families) who attend and do not attend preschool. However, states rarely have detailed, accurate data on child and family characteristics that can be used in such statistical models, and the characteristics that lead families to choose preschool education for their children may not be readily measured even in principle. The result is that estimates of program impacts are likely to be biased because the effects of the program are confounded with the effects of unmeasured child and family characteristics, leaving uncertainty about the effectiveness of contemporary preschool programs.

In 2004 the National Institute for Early Education Research (NIEER) partnered with state government and higher education in five states to employ an innovative, rigorous approach to evaluate state prekindergarten programs. This approach specifically addresses the selection bias problem and is applicable even when a prekindergarten program is truly universal. In each state

we used common methods and measures to assess the impact of the state programs for 4-year-olds on school readiness at kindergarten entry. Although substantial samples were obtained in each state, the study was designed so that the state samples could be pooled to increase the study's statistical power. The research design used is a regression-discontinuity (RD) approach which has a high degree of validity and corrects for selection bias. (Trochim, 1984). This method has been successfully used to evaluate Oklahoma's universal preschool program in Tulsa (Gormley, Gayer, Phillips, & Dawson, 2004).

The State Prekindergarten Programs

The present study was conducted across five states: Michigan, New Jersey, Oklahoma, South Carolina and West Virginia. The programs in Michigan, New Jersey and South Carolina target at-risk children while the programs in Oklahoma and West Virginia are universal. Each state program is unique, but all required licensed teachers with four-year college degrees and certification in early childhood. In Michigan, a small percentage of children attend private programs that do not have to meet the same standards for teacher qualifications as public schools. All programs serve children at age 4, though New Jersey's Abbott program serves nearly 80 percent as many children at age 3 as well. In New Jersey, we included only the state's "Abbott District" preschool program, the largest and best funded of that state's three preschool programs. Some states primarily provided services through the public schools, some primarily through private programs. All are well established, though New Jersey's Abbott program upgraded its standards for teachers and class size relatively recently. The chart below describes relevant characteristics of each state program. More detailed descriptions are available from Barnett et al. (2004).

Description of State Prekindergarten Programs Studied

State	Year established	Number served by child age	Percent of 4's enrolled	Length of day	Teacher/child ratio	Max. class size
Michigan	1985	24,729 age 4	19%	Half	1:8	18
New Jersey Abbott	1998	21,286 age 4 16,725 age 3	79% in Abbotts*	Full	2:15	15
Oklahoma	1990 –universal in 1998	30,180 age 4	65%	Varies	1:10	20
South Carolina	1984	17,821 age 4 740 age 3	32%	Half	1:10	20
West Virginia	1983 – universal by 2010	6,541 age 4 1,370 age 3	33%	Varies	1:10	20

* New Jersey's Abbott districts include about 1/4 of the state's children, statewide enrollment in Abbott and non-Abbott state pre-K was 25% at age 4.

Methods

Research Design

This evaluation employs a regression-discontinuity (RD) design, a statistical model with several strengths. The design reduces the likelihood of serious problems due to one of the most vexing problems in educational research, that of selection bias. Typically, program effects are estimated by comparing the test scores of children who attended a program with the scores of similar children who did not go. Where programs are universal, the problem of finding a “comparable” group of children who did not go to preschool is obvious. Yet, even where programs target only some children, a problem remains: those who go to preschool are *not* the same as those who do not. Preschool programs that target specific types of children create these differences through their eligibility criteria, but differences also come about because some parents choose to enroll their children and others do not. In sum, children who go to preschool differ from those who do not because programs select children and families select programs.

The RD design solution is to compare two groups of children who select, and are selected by, a state prekindergarten program, taking advantage of the stringent age (birth date) cutoff that states use in defining enrollment eligibility to define groups. This design is conceptually easier to understand by taking the extreme case: consider two children who differ only in that one was born the day before the age cutoff and the other the day after. When both are about to turn 5 years old the slightly younger child will enter the preschool program and the slightly older child will enter kindergarten having already attended the preschool program. If both are tested at that time, the difference in their scores can provide an unbiased estimate of the preschool program’s effect. Obviously, if only children with birthdays one day on either side of the age cutoff were included in a study, the sample size would be unreasonably small. However, the approach can be applied to wider age ranges around the cutoff. In fact, all children entering kindergarten from the state preschool program, and all children beginning preschool in the same year can be included in the study using RD statistical techniques.

Sample

Typically, samples of children who attend public school programs are drawn from roster lists provided by the school districts. However, we find that many school districts have a difficult time producing valid lists early in the school year, causing delay for research. Since the current research depends on assessing children as early as possible in the academic year, we developed a sampling strategy that required no student lists be provided. Our method was to gather information on the number and location of classrooms across the state universe of state-funded preschool and kindergarten programs and to randomly select enough state-funded preschool classrooms to provide us with the required preschool child sample, assuming approximately four randomly selected children per classroom. We then sampled the same number of kindergarten classrooms as preschool classrooms from each school district with preschool classrooms in the study. Trained research staff visited each sampled program site, selected children into the sample using the class roster and a random number list, and conducted the child assessments as early as possible in the school year.

We initially identified a random sample of 1937 classrooms (approximately half preschool and half kindergarten) in the five states, which would have yielded a sample of over 7600 children. Difficulties obtaining access to some classrooms (for example district refusals to allow participation based on passive consent) and scheduling problems led us to access 1320 classrooms, typically assessing four children per class. Thus, we collected data on 5278 preschool and kindergarten children in the five states. The preschool treatment group includes 2728 children, and the control group includes 2550 children. The sample is quite diverse: 47 percent were White, 25 percent African-American, 21 percent Hispanic, 3 percent Native American, and 2 percent Asian. The sample was roughly evenly split by gender, 48 percent were boys and 52 percent were girls.

Data Collection Procedures

In each state we worked with a local research partner to train child assessors on issues related to assessing children in school environments, confidentiality, protocol and professional etiquette as well as training specific to the assessment instruments and sampling procedures. Assessors were trained on each assessment and then shadow scored in practice assessments. Site coordinators were responsible for assuring adequate reliability throughout the study. A liaison at each site gathered information on the children's preschool status, usually from existing school records but occasionally from parent report, and was reimbursed \$5 per child for obtaining the information.

Children were tested in the fall of the 2004-05 school year. On all measures, children were tested in English or Spanish depending on their strongest language, which was ascertained from the classroom teacher. A very small number of children who did not speak either English or Spanish well enough to be tested were not included in the sample. Assessments were conducted one-on-one in the child's school, and assessments were scheduled to avoid meal, nap and outdoor play times. Testing sessions lasted 20-40 minutes.

Individualized assessments were selected to measure the contributions of the preschool programs to children's learning, with emphasis on skills important for early school success. Criteria for selection of measures included: (1) availability of equivalent tasks in Spanish and English, (2) reliability and validity, particularly pre-literacy skills that are good predictors of later reading ability; and (3) appropriateness for children ages 3 to 5. Each measure is discussed in detail below.

Measures of School Readiness

Children's receptive vocabulary was measured by the Peabody Picture Vocabulary Test, 3rd Edition (PPVT-3) (Dunn & Dunn, 1997). The PPVT – III is a 204-item test in standard English administered by having children point to one of four pictures shown when given a word to identify. The PPVT-III directly measures vocabulary size and the rank order of item difficulties is highly correlated with the frequency with which words are used. This test is also used as a quick indicator of general cognitive ability, and it correlates reasonably well with other measures of linguistic and cognitive development related to school success. Children tested in

Spanish were given the Test de Vocabulario en Imagenes Peabody (TVIP; Dunn, Lugo, Padilla, & Dunn, 1986). The TVIP uses 125 translated items from the PPVT to assess receptive vocabulary acquisition of Spanish-speaking and bilingual students.

The PPVT has been used for many years (over several versions) and substantial information is available on its technical properties. Reliability is good as judged by either split-half reliabilities or test-retest reliabilities. The test is adaptive in that the assessor establishes 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. This is important for avoiding floor and ceiling problems (Rock & Stenner, 2005). The PPVT-III and TVIP both have a mean standard score of 100 and a standard deviation of 15.

Children's early mathematical skills were measured with the Woodcock-Johnson Tests of Achievement, 3rd Edition (Woodcock, McGrew & Mather, 2001) Subtest 10 Applied Problems. Spanish-speakers were given the Bateria Woodcock-Munoz Pruebas de Aprovechamiento – Revisado (Woodcock & Munoz, 1990) Prueba 25, Problemas Aplicados. The manuals report good reliability for the Woodcock-Johnson achievement subtests, and they have been widely and successfully used in studies of the effects of preschool programs including Head Start. The achievement subtests have been standardized with a mean of 100 and a standard deviation of 15.

Phonological skills development was measured using the Blending subtest of the Preschool Comprehensive Test of Phonological & 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, the Pre-CTOPPP has been used with middle-income and low-income samples and includes a Spanish version. As the Pre-CTOPP has only been very recently developed, very little technical information is available about its performance and psychometric properties.

The Blending subtest includes items that measure whether children can blend initial phonemes onto one-syllable words, initial syllables onto two-syllable words, and ending phonemes onto one-syllable words. The percentage of items the child answered correctly out of 21 total subtest items is reported.

Print Awareness abilities were measured using the Print Awareness subtest of the Pre-CTOPPP. Items 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 36 total subtest items is reported. As noted above, this is a new instrument with little technical data.

Statistical Analyses

To estimate the effects of the prekindergarten programs on children's test scores we conducted a series of analyses that guard against bias due to model misspecification and gradually pruned the analyses to maximize the efficiency of the estimates. We used a single equation model for each outcome measure. The model accounted for the number of days

between birthdates and enrollment cut-off dates for each sample child, gender, ethnicity (classified as African-American, Hispanic, Native American, or White), and whether they were tested in English or Spanish. In addition, models were estimated with and without interactions for each state and treatment. Analyses were conducted using raw scores.

In the single equation model, the effect of attending the preschool program is estimated at the birth date cut-off for enrollment. A “treatment” variable was defined by assigning all children with birth date after cut-off date with a value of one (treatment) and all other children a value of zero (comparison). The selection variable (the age difference between birth date and cut-off date) was rescaled so that zero-point corresponded to the cut point. Thus, children in the treatment group had positive values, and children in the comparison group had negative values. An interaction term was constructed by multiplying the cut-off dummy variable by the rescaled selection variable. Interaction terms for state and treatment were used to investigate differences in program effects among states. As discussed below, such difference should not be interpreted as necessarily reflecting differences among state programs as they also can reflect differences in experiences of the comparison group.

As there is no *a priori* expectation that the estimated relationship should be linear, we estimated higher order polynomial forms of the equation, including squared and cubic transformations of the selection variable (the difference between birth date and cut-off date) and its interaction with the cut-off dummy variable). We began analyzing third-order (cubic) polynomial regression models and found the coefficients for cubic term (X^3) and its interaction with cut-off dummy variable (X^3Z) were rarely statistically significant. These two terms were dropped and the second order model was estimated. When we estimated the second order polynomial, the coefficients for the quadratic term (X^2) and quadratic interaction term (X^2Z) were not significant except for the quadratic interaction term in estimating children’s scores on the Print Awareness subtest. Thus, we dropped the quadratic term and its interaction term for estimating effects on the PPVT, Blending, and Applied Problems test scores. The point estimates for the effects of the programs were essentially unchanged across all of these alternative specifications.

For the regression discontinuity design to be effective, programs must adhere to a fairly strict use of a birth-date cut-off date for program enrollment to determine whether children are enrolled into the kindergarten or prekindergarten program based on their age. Each sample state employed a birth-date cut-off date for program enrollment, which varied by state. Children qualified to attend prekindergarten in academic year 2003-04 if they were born before September 1, 1999 in South Carolina, Oklahoma and West Virginia, or before December 1, 1999 in Michigan. They qualified to attend kindergarten if they were one year older, or born before those dates in 2000. In New Jersey, the age cut-off for program enrollment varied by school district from September 30 through December 31.

Our primary analyses were “sharp” regression-discontinuity models that employed a total 5071 children in our sample, dropping 207 children (4 percent of the total) whose birth date information appears to be inconsistent with the birth-date cut-off requirement for their programs. The 207 dropped includes both children who appeared to be too young for their grade ($n = 60$) and children who appeared to be too old for their grade ($n = 147$). “Fuzzy” analyses conducted

using an instrumental variables approach included these children and provided a way to see if their exclusion might have changed the results. Point estimates were virtually identical with the “sharp” analyses. Thus, results of the “sharp” analyses are reported here as they provide more statistical power than the instrumental variables analysis.

We also conducted analyses including a measure of whether or not each child qualified for free or reduced lunch under the federal subsidized school lunch program. Unfortunately, this information could not be obtained for 17 percent of the sample. The individual state response rates for this information varied from a particularly low 47 percent in one state to around 90 percent for the others. Analyses including dummy variables for free and reduced lunch and interactions between these and treatment are discussed only briefly as they produced essentially the same results as the primary analyses.

One of the key assumptions underlying the regression discontinuity design is that the unobservable characteristics of children do not vary discontinuously around the birth date cutoff. Obviously, this is not directly testable. However, it is possible to see if the observable characteristics vary discontinuously at the cutoff, which is at least suggestive of a possible problem. To test this we re-estimated the regression discontinuity model with minority status and free lunch status as dependent variables. These analyses did not find statistically significant discontinuities at the age cutoff. Details of these and other secondary analyses are available upon request.

Results

The main results for the estimated effects of state-funded preschool on children’s scores are displayed in individual figures for each outcome measure. Each figure displays a regression line of the children’s predicted test scores by the distance in days their birth date is from the program enrollment cut-off date. The discontinuity in the regression line at the cut-off date is the estimated effect of the preschool program.

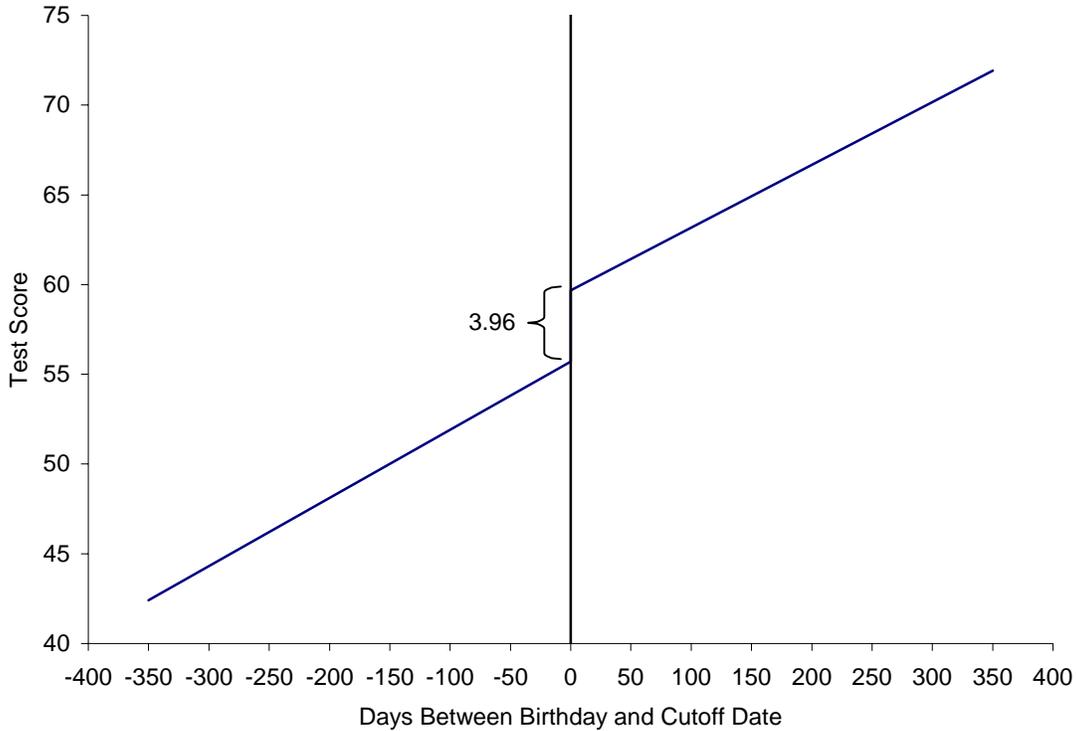
Results are also summarized in Table 1, which presents descriptive statistics for the sample, and in Table 2, which presents the estimated equations including the average effects on children’s test scores across the five state programs. Statistically significant effects of the state preschool programs were found on four of the five outcome measures for the sample as a whole and for each state. Results for each outcome measure are reviewed in detail below.

Vocabulary

The estimated effect of state-funded prekindergarten on children’s receptive vocabulary was statistically significant across the total sample. Individual state estimates did not significantly differ from each other. State-funded preschool improved children’s scores by 3.96 raw score points, which amounts to 31 percent more growth in vocabulary scores over the year due to the program and an 8% increase in average scores. Across all states this improvement is about 21 percent of a raw score standard deviation. This improvement corresponds to about three months of vocabulary development. A gain of four raw score points translates into a gain of about four standard score points, or 26 percent of a normed standard deviation.

Figure 1 below portrays a regression line of the children’s predicted PPVT scores by the distance in days their birth-date is from the program enrollment cut-off date. The discontinuity in the regression line at the cut-off date represents the estimated effect of the preschool program, or 3.96 raw score points.

Figure 1. Prekindergarten Education’s Effects on Receptive Vocabulary

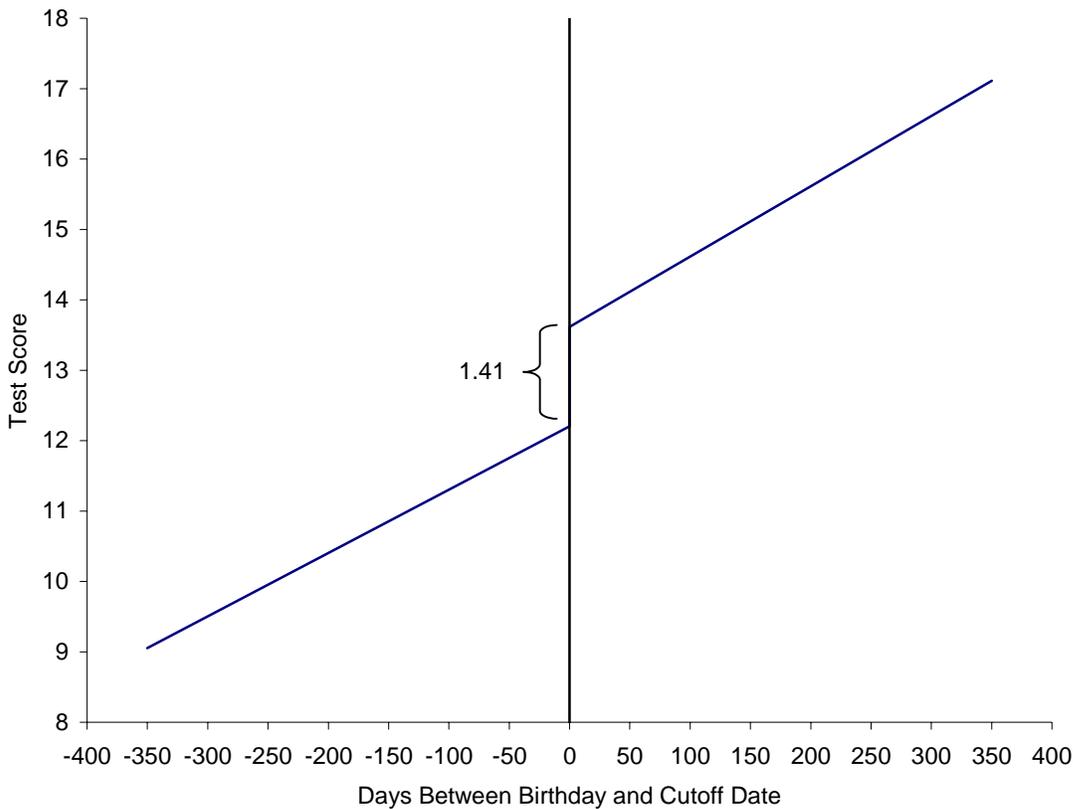


Math

The effect of state-funded prekindergarten on children’s early math skills as measured by the Woodcock-Johnson-III Applied Problems subtest is statistically significant for the overall study and for individual state programs. Results indicate that prekindergarten improves children’s math scores by about 1.41 raw score points overall, averaged across four states (math was not assessed in South Carolina, as this measure was added to the study after work began in South Carolina), or 35 percent of a raw score standard deviation. The estimated effect is significantly greater for some states than others, by up to 1.3 raw score points. In this age range, one raw score point is equivalent to about three standard score points. Thus, average improvement across the four states translates into about 4.2 standard score points or 28 percent of a normed standard deviation. This amounts to 44 percent more growth over a year due to the program and a 13 percent increase in average scores.

Figure 2 below portrays a regression line of the children’s predicted Applied Problems scores by the distance in days their birth date is from the program enrollment cut-off date. The discontinuity in the regression line at the cut-off date represents the estimated effect of the preschool program, or 1.41 raw score points.

Figure 2. Prekindergarten Education’s Effect on Math Skills

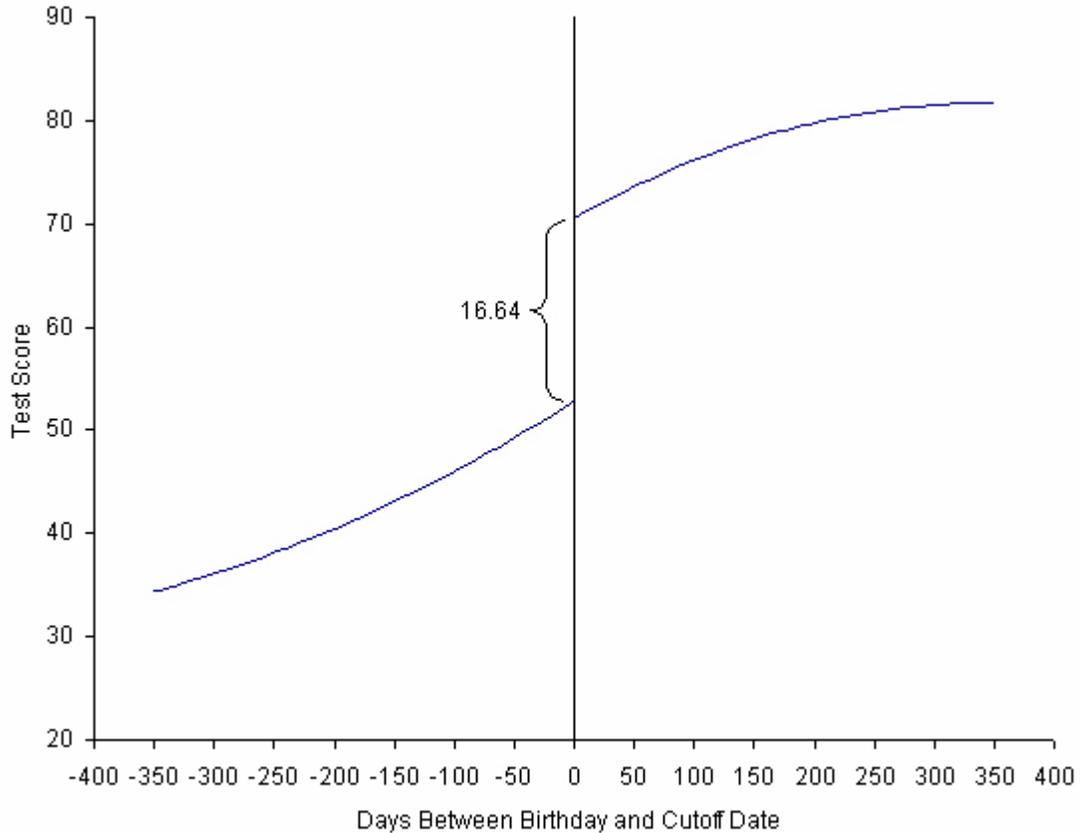


Print Awareness

The estimated effect of state-funded prekindergarten on children’s Print Awareness scores was statistically significant overall and for each state program. The estimated effect is about 16.64 percent more items answered correctly (averaged across the five states), or about 64 percent of a standard deviation. Some states had significantly greater effects than others, by up to almost 11 percent more items answered correctly. Another perspective on the average effect is that it amounts to 85 percent more growth in print awareness over the year due to the program and a 39% increase in children’s average print awareness scores.

Figure 3 below portrays a regression line of the children’s predicted Print Awareness scores by the distance in days their birth date is from the program enrollment cut-off date. The discontinuity in the regression line at the cut-off date represents the estimated effect of the preschool program, or 16.64 raw score points.

Figure 3. Prekindergarten Education’s Effects on Print Awareness



Phonological Awareness

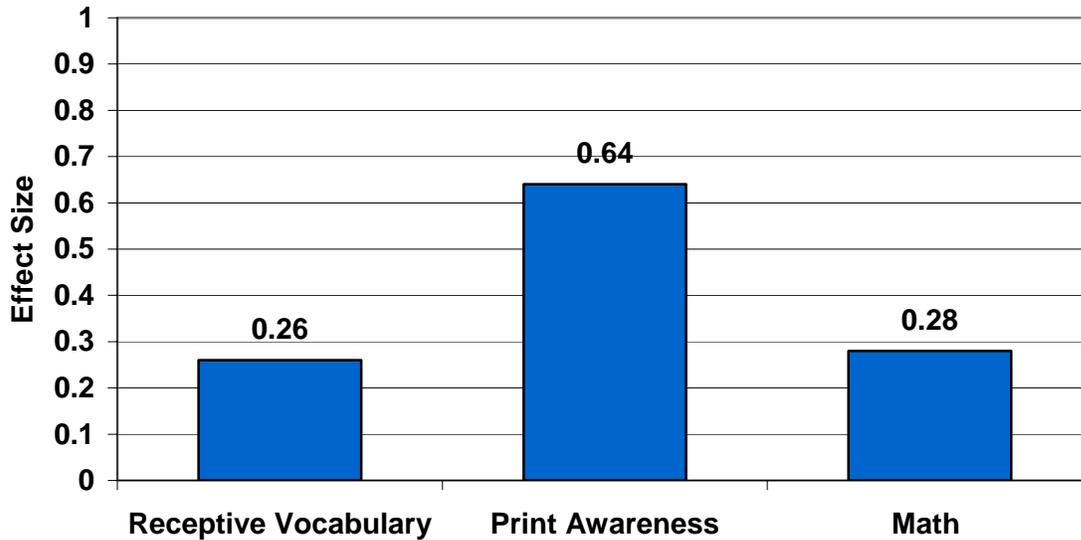
No statistically significant effect of prekindergarten was found on phonological awareness as measured by the Blending subtest, either overall or for any of the individual states. On all of the other measures children in this study scored below the reported averages. This is hardly surprising as most of the programs targeted disadvantaged children. However, the overall sample and even the highly disadvantaged New Jersey Abbott District sample scored above the averages reported for the Blending subtest at entry to preschool. The sample as a whole had a mean of 14.24 at age 4 and 16.29 at age 5 compared to means reported by the test authors of 12.47 at age 4 and 13.75 at age 5 (Pre-CTOPPP Website, 2002).

Summary

By way of summary, Figure 4 below portrays the effect sizes of the impact of state-funded preschool on children’s receptive vocabulary, print awareness and math scores. These

effect sizes are another way of standardizing the estimated effects so that they may be compared to estimated effects in other studies.

Figure 4. The Effect of State-Funded Preschool on Children's Scores across Measures



Prekindergarten Effects and Family Income

Family income, measured whether children qualify for free lunch, reduced price lunch or neither (as reported by the school), was not included in the primary analyses presented here because missing data on this measure reduced sample size by nearly 20 percent overall and by more than 50 percent in one state. However, additional analyses were conducted to test whether children from lower income families who qualified for free and reduced price lunch benefited more than those who did not qualify because they were from somewhat higher income families. There is some evidence for a stronger effect of the program on print awareness skills for children from lower income families. This effect approaches significance in the larger study. Overall, children gains from the preschool program were about 3 more items correct for those who qualify for free or reduced price lunch compared to gains for children who did not qualify. In the two states with statistically significant differences in effects by income (Oklahoma and South Carolina) the extra gain from pre-k for children from lower income families was about 8 percent more items correct. These were the only significant differences we found in effects by income. Otherwise, results are virtually identical to those presented earlier when free lunch status is included in the analyses.

Conclusion

This study of state prekindergarten programs serving 4-year-olds estimated the effects on children's learning in five states that represent a broad cross section of the United States--north and south, east and west, urban and rural. Remarkably, across them all we found significant and

meaningful effects on children's language, literacy, and math skills. This study provides strong evidence that quality preschool programs produce broad gains in children's learning and development at kindergarten entry. These preschool programs produce the kinds of effects that can be expected to yield greater school success and later improvements in reading and math. For example, early print awareness and receptive vocabulary have been found to predict later reading abilities in the early elementary grades (Snow, Burns, & Griffin, 1998). The effects found in this study are the first link in a chain that produces the long-term school success and economic benefits documented by preschool studies that have followed children into adulthood (Schweinhart, Montie, Ziang, Barnett, Belfield, & Nores, 2005; Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002; Reynolds, Temple, Robertson, & Mann, 2002).

This study's results are consistent with findings from other rigorous studies of state preschool education programs (Gormley et al., 2004; Barnett et al., 2004; Frede & Barnett, 1992; Irvine, Horan, Flint, Kukuk, & Hick, 1982). Where direct comparisons can be made, the size of the program impacts is quite similar to those found in the recent study of Oklahoma's program in Tulsa. These estimated effects for state-funded prekindergarten programs are smaller than those found for highly intensive model programs that had much better student-teacher ratios and provided more than one-year of education at age 4 (Barnett, 1998), and are larger than those found in the recent national impact study of the federal Head Start program.

The use of identical or similar measures in the Head Start Impact Study permits direct comparisons of findings with this study (Puma, Bell, Cook, Heid, Lopez, Zill, et al., 2005). The Head Start National Impact study found no statistically significant effects for 4-year-olds on receptive vocabulary (PPVT) or early math scores (WJ-III Applied Problems). Our study finds significant effects of state-funded preschool programs on these outcome measures. Researchers use standardized "effect sizes" to compare effects across studies. The National Impact study does not report these for Head Start's effects on the PPVT and Applied Problems tests are not reported, but it can be deduced from the Impact Study that they are no more than about 0.10 scores. By comparison, the effects that we find for state prekindergarten are 0.21 for the PPVT and 0.35 for the applied problems. The largest effect size in our study was for print awareness and equaled 0.61; this could be compared to the 0.22-0.24 effect sizes for letter naming and letter-word identification in the Head Start impact study. Thus, the effects we find are at least 2 to 3 times as large as in the Head Start study. This difference in outcomes between the two types of programs points to the likely effects of the higher qualifications (and higher compensation) of teachers in state prekindergarten programs compared to Head Start. The state prekindergarten programs we studied do not uniformly differ from Head Start with respect to other characteristics such as length of day or class size.

We did not find that state-funded preschool programs significantly improved children's blending skills, our sole measure of phonological awareness. (It is notable that the Head Start impact study also found no significant effects on phonological awareness at any age.) There are several plausible explanations for this finding. Perhaps these preschool classrooms did not provide as much support for these skills as they did for language development and print awareness (Lamy & Frede, 2005). In that case, activities and interactions to support children's phonological sensitivity – hearing smaller sounds within the spoken word that may be parsed out and switched for others to create rhymes and alternate endings – may need to be increased. However, caution is warranted since there are plausible alternative explanations for this finding.

As we reported, the children in this study were already doing well on the phonological awareness measure compared to the averages reported by the measure's authors. We cannot rule out that children were doing so well with these skills in the absence of preschool education that prekindergarten had little opportunity to improve their skills. However, the fact that even highly disadvantaged children had relatively high scores on the measure (when they had much lower relative scores on the other measures) raises questions about the measure itself. It simply may not measure these skills well in children at ages 4 and 5. We adopted the measure because other alternatives are scarce (in Spanish and English), but as a new measure it lacks data that would establish its validity. Our results suggest that more research is needed on the measure itself, as well as on the extent to which preschool classrooms might need to increase their emphasis on phonological development.

The size of the estimated effects of prekindergarten varied somewhat among the states on print awareness and math skills, but not on vocabulary or phonological awareness. It is possible that some of the variations are due to differences among the programs. However, the "control" condition is not the same across states, nor is population served. The samples varied across states, from highly disadvantaged to a cross-section of the general population, just as programs vary from highly targeted to universal. The educational circumstances of the comparison group varied across states, as well. Thus, our study does not provide a particularly strong basis for seeking to explain why the estimated effects vary among state preschool programs that are similar with respect to such structural parameters as teacher qualifications and class size. Further research to address this question would benefit from more information on the content and practices of each state's preschool programs and the settings that children would attend in lieu of a state prekindergarten program.

It is important to understand that the design we employed does not estimate the effect of the state prekindergarten programs compared to no preschool education. Our approach estimates the effect of a program at age four compared to whatever alternatives are obtained by the children's families. For example, in New Jersey's Abbott District program, the vast majority of the comparison group children attended a state-funded preschool program at age three. Thus, we essentially estimated the effect of adding a second year of preschool education in New Jersey. In the other states, many fewer children (in some states none) attend the state prekindergarten program at age 3, but children could attend Head Start or a private preschool program or child care center at age 3 and it is likely that many did. (Yarosz & Barnett, 2001).

In sum, this study provides rigorous evidence that a broad cross section of quality state prekindergarten programs for 4-year-olds produce substantial gains in children's learning and development. While not exactly the same, these state programs all share a commitment to high standards for teacher qualifications and adequate compensation. The effects are equally strong for programs that target disadvantaged children and for those that seek to serve all children. Looking within as well as across states we find some evidence that even though all children benefit from pre-k, the benefits may be somewhat larger for children from lower income families.

Comparisons to Head Start suggest that public preschool programs with weak standards for teacher qualifications (and low teacher pay) might increase their effectiveness by raising their teacher qualifications standards and compensating teachers accordingly. Of course, even the programs studied here should not be complacent. Studies of the effects of model programs for disadvantaged children indicate that even more can be done, at least for those children most in need. Even state prekindergarten programs with high standards should monitor and evaluate their performance with a view toward continuous improvement so that preschool education can be even more effective in the future.

Table 1. Children’s Demographics and Scores by Group

Demographics	No Prek	Prek
Ethnicity		
African American	27%	24%
White	47%	8%
Hispanic	20%	23%
American Indian	3%	3%
Asian	2%	2%
Other	2%	2%
Free lunch	55%	55%
Gender		
Boys	48%	49%
Girls	52%	51%
Home language English	83%	84%
Tested in Spanish	3%	2%
Scores		
Receptive language		
Raw score	49.02 (18.87)	65.68 (18.58)
Standard score	92.16 (15.38)	94.10 (14.37)
Math		
Raw scores	10.66 (4.07)	15.31 (4.31)
Standard score	97.08 (14.15)	95.89 (12.94)
Print Awareness (% correct)	43.48 (25.83)	78.70 (21.47)
Phonological Awareness (% correct)	67.94 (23.70)	77.72 (22.44)

Table 2. The Effects of the Preschool on Test Scores: Regression Coefficient and Standard Errors

Independent Variables	Dependent Variables			
	PPVT	Blending (%)	Print Awareness (%)	Applied Problems
Preschool effect	3.964*** (.881)	-.145 (1.283)	16.641***(1.955)	1.409***(.236)
Qualify (days)	.038*** (.003)	.027***(.004)	.079***(.018)	.009***(.001)
Qualify ²	-----	-----	6.66E-005 (.000)	-----
Qualify* Cut-Off	-.006 (.004)	.002 (.006)	-.012 (.024)	.001 (.001)
Qualify ² * Cut-Off	-----	-----	.000** (.000)	-----
Black	-8.243*** (.581)	-3.525***(.833)	-1.064 (.865)	-1.278***(.164)
Hispanic	-13.729*** (.684)	-6.082***(.994)	-9.311***(1.030)	-1.807***(.172)
Native American	-6.048***(1.548)	-.725 (2.185)	-6.207** (2.275)	-1.499***(.375)
Girls	1.652*** (.434)	3.220***(.630)	5.627***(.649)	.511***(.116)
English/Spanish test	22.963***(1.459)	14.264***(2.358)	10.240***(2.463)	2.287***(.358)

Notes: Standard errors in parentheses. * $p < .05$, ** $p < .01$, *** $p < .001$

References

- Barnett, W.S. (1998). Long-term effects on cognitive development and school success. In W.S. Barnett & S.S. Boocock (Eds.) *Early care and education for children in poverty*, (pp. 11-44). Albany: SUNY 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.
- Barnett, W.S., Robin, K.B., Hustedt, J.T., & Schulman, K.L. (2004). *The state of preschool: 2004 state preschool yearbook*. New Brunswick, NJ: National Institute for Early Education Research.
- 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.
- Dunn, L. M. & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test-Third Edition (PPVT-3)*. Circle Pines, MN: AGS Publishing.
- Dunn, Padilla, Lugo & Dunn, (1986). *Test de Vocabulario en Imágenes Peabody (TVIP)*. Circle Pines, MN: AGS Publishing.
- Frede, E. & Barnett, W.S. (1992). Developmentally appropriate public school preschool: A study of implementation of the High/Scope curriculum and its effects on disadvantaged children's skills at first grade. *Early Childhood Research Quarterly*, 7, 483-499.
- Gormley, W., Gayer, T., Phillips, D., & Dawson, B. (2004). *The effects of universal pre-k on cognitive development* (Working paper #4). Washington, DC: Georgetown University.
- Gresham, F. M. & Elliot, S. N. (1990). *Social Skills Rating System (SSRS) – Teacher form*. Circle Pines, MN: AGS Publishing.
- Irvine, D. J., Horan, M. D., Flint, D. L., Kukuk, S. E. & Hick, T. L. (1982). Evidence supporting comprehensive early childhood education for disadvantaged children. *Annals of the American Academy of Political and Social Science*, 461(May), 74-80.
- Lamy, C. & Frede, E.C. (2005). *Giant steps for the littlest ones: progress in the sixth year of the Abbott preschool program*. New Jersey Department of Education, Office of Early Childhood Education, Trenton, NJ.
- Lonigan, C., Wagner, R., Torgeson, J. & Rashotte, C. (2002). Preschool Comprehensive Test of Phonological & Print Processing (Pre-CTOPPP). Department of Psychology, Florida State University.

- Pre-CTOPPP Website (2002). Preschool Comprehensive Test of Phonological and Print Processing [Data File]. *Subtest statistics for preschool comprehensive test of phonological and print processing by age group*. Available from Lonigan's Website at <http://www.psy.fsu.edu/~lonigan/data.pdf>
- Puma, M., Bell, S., Cook, R., Heid, C., Lopez, M., Zill, N., Shapiro, G., Broene, P., Mekos, D., Rohacek, M., Quinn, L., Adams, G., Freidman, J. & Bernstein, H. (2005). *Head Start impact study: First year findings*. Washington, DC: US Department of Health and Human Services, Administration for Children and Families.
- 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 on line at <http://www.ssc.wisc.edu/irp/pubs/dp124502.pdf>.
- Rock, D.A., & Stenner, J.A. (2005). Assessment issues in the testing of children at school entry. *Future of Children*, 15 (1), 15-34.
- 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.
- Snow, C., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Trochim, W. M. K. (1984). *Research design for program evaluation: The regression-discontinuity approach*. Beverly Hills, CA: Sage Publications.
- Wagner, R., Torgeson, J. & Rashotte, C. (1999). *Comprehensive Test of Phonological Processing (CTOPP)*. Austin, TX: Pro-Ed.
- 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.
- Yarosz, D. & Barnett, W.S. (2001). Who reads to young children? Identifying predictors of family reading activities. *Reading Psychology*, 22(1), 67-81.