Journal of Education for Students Placed at Risk, 18: 128–140, 2013

ISSN: 1082-4669 print/1532-7671 online
DOI: 10.1080/10824669.2013.799424

Getting Ready for School: Piloting Universal Prekindergarten in an Urban County

Robert L. Fischer

Mandel School of Applied Social Sciences, Case Western Reserve University

Lance T. Peterson

University of St. Thomas/St. Catherine University, School of Social Work

Tirth Bhatta

Department of Sociology, Case Western Reserve University

Claudia Coulton

Mandel School of Applied Social Sciences, Case Western Reserve University

Investments in high-quality early care and education have been shown to reap societal benefits across the lives of the children served. A key intervention point is in the lives of 3- to 5-year olds during the period prior to entering kindergarten. Many jurisdictions have developed broad-based prekindergarten initiatives. This study reports on a pilot universal prekindergarten program in 24 sites in the Cleveland, Ohio area. Child assessment data were collected on 204 children from early care classrooms for 3- to 5-year olds across 3 time points by trained observers using 2 standardized instruments. Changes in achievement scores were shown to be significantly predicted by race, parental education level, and whether the family spoke English as a second language, with largest gains shown among children who were most behind at baseline. The findings serve to illuminate the developmental trajectory of children before kindergarten and how data can be used to inform practice and policy.

The ability of young children to arrive in kindergarten ready to learn is a major concern among communities in the country and throughout the world. Underlying this concern is a growing knowledge about children's rapid brain development during the first 3 years of life (Karoly et al., 1998), as well as the greater risk of school failure if children reach school age with developmental deficits (Olsen & DeBoise, 2007). Accordingly, several early-childcare and education programs have been created in recent years across the country to promote the school readiness of children (Brown & Scott-Little, 2003).

Correspondence should be addressed to Robert L. Fischer, Ph.D., Mandel School of Applied Social Sciences, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106-7164. E-mail: fischer@case.edu

EARLY CHILDHOOD EDUCATION AND CHILDREN'S ACHIEVEMENT

Two reviews of research on children's achievement (i.e., progress in language, math skills, or both) show that early childhood education and child care have positive benefits for the development of language and mathematical skills. Brown and Scott-Little (2003) conducted a review of 20 studies assessing the school readiness of children who participated in early childhood education (i.e., programs that potentially targeted children from birth to age 5). The authors of three out of the four experimental or quasi-experimental studies reported significant results for language and literacy skills. Barnett (1995) found that of 11 model early childcare and education (ECE) programs, five reported significant positive effects of ECE on achievement beyond the third grade; one study even found achievement effects into junior high school. Collectively, these reviews illustrate that although findings are mixed related to children's achievement, evidence suggests that children's achievement improves as a consequence of early childhood education.

More recent examples also suggest that early childhood education may boost child achievement levels. Abbott-Shim, Lambert, and McCarty (2003) found that children who participated in Head Start showed a significantly faster rate of growth on receptive vocabulary and phonemic awareness than a comparison group. Gormley and Phillips (2005) found that overall test scores for all children exposed to the Tulsa Public Schools prekindergarten program increased 16% on average, primarily due to improvements on language and cognitive skills. Hispanic children's test scores improved by 54%; African American children's test scores improved by 17%.

UNIVERSAL PREKINDERGARTEN (UPK) PROGRAMS

Given the pattern of benefits of early childhood education, several US states have adopted UPK programs to increase opportunities for young children. UPK programs are characterized by preschool being available for "all four-year-olds regardless of income or other identifiable risk factors" (Brown & Scott-Little, 2003, p. 5). Specifically, seven states, Florida, Georgia, New York, Oklahoma, New Jersey, Wisconsin, and West Virginia (Kirp, 2005; Schumacher, Ewen, & Hart, 2005), as well as the District of Columbia (Brown & Scott-Little, 2003), have initiated or are working toward a UPK program. One recent UPK program is in Oklahoma, which uses its public school system to deliver UPK services (Gormley and Phillips, 2005).

Outcomes reported from the Georgia and Oklahoma preschool programs show promise. Henry et al. (2003) reported on an assessment of the Georgia Prekindergarten Program, which included comparisons to private preschools and Head Start. Children were assessed at the beginning of preschool and again at the beginning of kindergarten. Henry et al. reported that on five assessments of achievement, after accounting for family risks and individual characteristics, children in the program entered kindergarten at similar levels of school readiness to private preschoolers and ahead of children attending Head Start. In Oklahoma, overall test scores for all children exposed to the Tulsa Public Schools UPK program increased 16% on average, primarily due to improvements on language and cognitive skills (Gormley & Phillips, 2005).

FACTORS ASSOCIATED WITH EARLY CHILDHOOD EDUCATION

One important factor related to children making gains in achievement is that children who start at the lowest levels make the largest gains over time. Such has been a finding of the Family and Child Experiences Survey (FACES), in which two nationally representative cohorts of 3- and 4-year old children in Head Start (1997, 2000) were evaluated (Zill, Resnick, Kim, O'Donnell, & Sorongon, 2003). In both cohorts, children who started with the lowest percentile rankings compared to national averages made the greatest gains over time in vocabulary, early writing skills, letter and word recognition, and early math skills.

Another important factor of early childhood education is the quality of care. Brown and Scott-Little (2003) discussed common themes associated with higher quality UPK programs, which include reasonable staff-to-child ratios and group sizes; specialized staff training in early childhood education; caring, responsive teachers; and a curriculum with a variety of developmental activities. Successful early childhood education programs tend to have these program characteristics (McCall, Larsen, & Ingram, 2003). More specifically, Bryant et al. (2003) found that overall classroom quality has been positively correlated with children's cognitive, language, social, and emotional outcomes.

THE INVEST IN CHILDREN INITIATIVE OF CUYAHOGA COUNTY

The Invest in Children (IIC) initiative in Cuyahoga County, Ohio, originally launched in 1999 as the Early Childhood Initiative, was created in recognition of the needs of young children. This comprehensive, bold initiative is a community-wide, public–private partnership of individuals and organizations including government agencies, community-based service providers, medical institutions, and philanthropic and private groups all working together to help increase the development, funding, visibility, and impact of early childhood services. An important component of IIC has been preparing children for school, which spawned a commitment to UPK. IIC partners convened a community-wide UPK planning effort in 2006. Over time, this led to studies of child-care capacity and quality, which were used to understand the early care and education context in Cuyahoga County and plan a UPK pilot study.

The purpose of this study is to report findings related to the kindergarten readiness of the children served through Cuyahoga County's UPK pilot program. Two questions are addressed in this study. First, do verbal and receptive language skills, as well as math and logical problem-solving skills, of children enrolled in UPK improve over time, beyond what would be expected through maturation? Second, do verbal and receptive language skills, as well as math and logical problem-solving skills, of children enrolled in UPK improve as a function of the children's initial percentile rank, the quality of the classroom, or both?

METHOD

Sample and Procedure

A two-stage sample selection procedure was used. First, a sample of 40 classrooms was randomly selected from the 61 classrooms participating in the pilot. Researchers received student

rosters organized by classroom which were used as the basis of random selection using SPSS statistical software. Second, for each selected classroom, a minimum of five children were randomly selected for invitation into the study. The names of potential study participants were provided to researchers by Cuyahoga County's Office of Early Childhood. The parents of selected children were contacted by phone about participating in the research and evaluation study and informed consent was secured. Participants were contacted and accepted until over 200 children had been recruited into the study. The final sample consisted of 208 children from 24 different UPK sites. Parents received three gift certificates in the amount of $25 each for allowing their children to participate in the three observations ($75) and an additional $25 for completing a parent survey. Children were given coloring books, stickers, and similar items. The study was conducted under an approved protocol from the University Institutional Review Board.

UPK children were assessed on measures of cognitive development at three time points: spring 2008, fall 2008, and spring 2009. Four interviewers were trained in assessment procedures and mandatory refresher courses were provided prior to phases two and three. All four conducted assessments during the first phase of the study (spring 2008). One of the four interviewers was unavailable for the remaining phases of the study; consequently, the three remaining interviewers conducted assessments in phases two (fall 2008) and three (spring 2009). Each interviewer was paired at least one time with another interviewer to ensure reliability of testing administration.

Given the importance of providing high-quality services, all UPK sites were required to meet specific criteria. These criteria closely align with the characteristics discussed by Brown and Scott-Little (2003) and revolve around teacher–child ratios, credentialing requirements of administrators and staff; accreditation of, and certification of, family child care providers, or both; research-based curricula; and the need to implement assessment tools relevant to measuring program effectiveness.

Participation in the pilot provided sites with supplemental funding and resources to meet and maintain adherence to these criteria. A core emphasis of the UPK pilot was to invest in the quality of care in participating sites to enhance the child outcomes for the children in care. The funding provided to UPK sites could be used to enhance teacher compensation, provide additional program supports, and acquire specific program resources and materials. To assess the level of quality in the center-based sites, a standardized measure of structural quality of care was administered mid-program-year in 2008 and 2009. The scores on the Early Childhood Environment Rating Scale—Revised (Harms, Clifford, & Cryer, 1998) are reported on a 7-point scale, including 1 (*inadequate*), 3 (*minimal*), 5 (*good*), and 7 (*excellent*). The program sites showed statistically significant ($p < .05$) improvements on all subscales and the total ECERS score between spring 2008 and spring 2009. The overall ECERS score improved by 18% and the largest gains were found in the areas of personal care routines (30% increase) and activities (35% increase). These gains are substantial for a 1-year period of investment in these center-based programs.

Measures

Child-level achievement was assessed in the UPK sample by means of two instruments. First, the Peabody Picture Vocabulary Test (PPVT-IV) was used to measure receptive language skills. It has been standardized, using norm samples for each age and grade level starting with age 2 years 6 months. Moreover, norm samples are reported as strongly representative of the US population.

Internal consistency for the age ranges represented in the UPK pilot study ranged between .95 and .97 in the norm samples. Other psychometric properties (e.g., test–retest reliability, convergent validity) support the use of the PPVT-IV for receptive language skills (Dunn & Dunn, 2007). Second, two subtests of the Woodcock Johnson-III test of cognitive ability (WJ-III) were used. The letter/word recognition subtest measures the child's ability to recognize words and letters and the applied problems subtest measures children's beginning math skills. The WJ-III has also been standardized using norm samples. One year test–retest reliability for the age 4–7 norm sample was $r = .92$ for both the WJ-LW and WJ-AP. Other tests of reliability and validity suggest that the WJ-III is an instrument with strong psychometric properties (Woodcock, McGrew, Schrank, & Mather, 2001, 2007).

A parent survey (not reported in this analysis) was used to ascertain parents' overall perspective on their children's experience in preschool. Several Likert-scale items were used to assess parents' ratings of the curriculum emphasis, as well as what they would have liked their children to receive from their preschool experience. Items also assessed parents' experience with managing the costs associated with preschool.

Analytical Plan

The analysis of the student-level data was undertaken in two phases. First, student scores were examined over time in the aggregate. Second, a multilevel model was applied to accommodate the nested nature of the student data. Because the test scores were standardized by age, the time point was treated as a continuous scale to estimate change in test scores across three time points. For the purpose of the analysis, the initial time point was coded as 0 and subsequent time points were coded as 1 and 2.

Several variables were used to represent the second level of characteristics at the child level. These were parental education, race, gender, and family size. Parental education, which ranged from grade 9 to some college, was treated as a continuous variable. Grade 9 was recoded as 1 and subsequent education levels were recoded in increasing order. For the purposes of our analysis, education was mean-centered to make the estimates more interpretable. Gender and race were employed as dummy variables. For gender, female was coded as 1 and male was coded as a 0. In terms of race, each racial group was dummy-coded and White was treated as the reference group for the purpose of the analysis. The number of people in the child's family was recoded into two categories. A child with fewer than five family members was coded as 1, whereas a child having more than five family members was assigned a score of 0. To examine whether parental language background matters for the progress of a child in prekindergarten level, one dummy variable was created to represent their primary language. A score of 1 was assigned to a child for whom English is not the native language, and a score of 0 was assigned to those for whom English is their native language.

The single level-3 variable was a measure of structural quality for the early care program, based on the State of Ohio's voluntary quality rating system (Step Up to Quality). Early-care programs apply to the rating system and are assigned 0–3 stars to denote their level of quality, where a rating of 3 indicates the highest quality level. The star rating is based on the site's characteristics in regard to child–teacher ratios, staff education and qualifications, staff receipt of specialized training, and use of evidence-based models and other approaches that promote early

TABLE 1
Multilevel Modeling Analysis Plan

At Level 1:	*At Level 2:*	*At Level 3:*
$Y_{tij} = \pi_{0ij} + \pi_{1ij}(\text{Time})_{tij} + e_{tij;}$	$\pi_{0ij} = \beta_{00j} + r_{0ij}$	$\beta_{00j} = \gamma_{000} + u_{00j}$
	$\pi_{1ij} = \beta_{10j} + r_{1ij}$	$\beta_{10j} = \gamma_{100} + u_{10j}$
Where:	Where:	
Y_{tij} = Test score at time t for child i at site j	β_{00j} = Mean initial test score within site j	
π_{0ij} = Initial test score of child i at site j	γ_{000} = Overall mean initial status	
π_{1ij} = Change in the test score of child i at site j during each wave	β_{10j} = Mean change in test score within site j	
	γ_{100} = Overall mean change in test score	

learning. For the purpose of the analysis, it was treated as a dummy variable, with the sites having a star rating of 3 coded as 1 and the sites having a star rating of less than 3 coded as 0.

The sampling design that was used to collect data lends itself well to multilevel analysis. The individual level samples were randomly drawn from each site and measured at three different time points. The potential for variation in the initial status and also in the change in the test score within each site and across different sites makes these data well suited for three-level analyses. The procedure, identified in Table 1, was adopted to set up a three-level model to predict initial status and change in the various test scores. Initially, an unconditional model was set up in which no predictors were included in the second and third levels. For example, level 1 included only time as a main independent variable to examine the change pattern across time of children *i* in site *j*.

No predictors were added at the second level or at the third level. The purpose of estimating this model was to treat it as a standard model to which the final model could be compared, to examine the influence of predictors at the second and third level on initial status and change in test scores. For the final model, the individual-level predictors were entered at the second level to assess their impact on the initial test score and the change in the test score. Moreover, the three-level variables were entered to examine whether their inclusion explains the variability in mean initial test score and mean change in test score within each site.

The descriptive analysis of the test scores showed that 14% of cases were missing at wave 2, primarily due to children changing settings. The multiple imputation procedure was performed to explore whether it makes any substantive difference to the model estimates. The model estimates obtained from the imputed data showed no significant differences on the model estimates, suggesting that complete case analysis provides fairly unbiased estimates. The results presented in this article are from complete case analysis. Statistical software, SAS, was used to perform all analyses.

Results

Child achievement was first examined over time and paired *t*-tests were conducted between achievement scores across time points. The analysis focused on the change between first and third time points, representing approximately 1 year. Results were computed based on the change in scores according to the child's initial percentile rank. Initial percentile rank was determined by how the child's standard score compared to the average standard score on each of the measures.

Achievement scores improved on all tests for those who started the study in the 50th percentile or below. Research has shown that a child having a single year of preschool, compared to no preschool, is associated with a standard score gain of 0.9 on the PPVT and gain of 2.5 on the WJ-AP (Barnett & Lamy, 2006). Results from other UPK programs, such as Georgia, show that gains on the PPVT average 4.7 (with the most disadvantaged children having gains of 7.0), and gains on the WJ-AP average 3.8 over 1 year (Henry et al., 2003).

Descriptive statistics for variables used in this study are presented in Table 2. The average age at Time 1 of the study participants is 3.6 years and 51% of them were girls. Significantly high numbers of participants were African American, comprising 55% of total participants. Seventy percent of participants' parents reported having higher than high school education. Seven percent reported that English was not their native language. Seventy-four percent reported having fewer than five members in a family. Also evident in Table 2 is that mean test scores increased for each achievement test across all three waves.

Table 3 presents the fixed and random estimates of a three-level unconditional model. The fixed model shows a statistically significant average increase in achievement test scores across

TABLE 2
Descriptive Statistics for Study Variables

Variable Name	*Value Range*	*Mean*	*SD*
Dependent Variables			
PPVT-IV Wave 1	56 to 149	97.73	14.73
PPVT-IV Wave 2	45 to 139	99.20	13.99
PPVT-IV Wave 3	70 to 142	100.21	12.50
WJ-III-LW Wave 1	69 to 155	102.55	12.81
WJ-III-LW Wave 2	68 to 153	103.31	13.07
WJ-III-LW Wave 3	70 to 149	104.96	13.20
WJ-III-AP Wave 1	65 to 140	101.82	12.61
WJ-III-AP Wave 2	60 to 141	102.40	12.90
WJ-III-AP Wave 3	70 to 156	103.84	12.80
Demographic characteristics			
Age (years)	3 to 5	3.60	.50
Female	1 = Female; 0 = Male	51%	
Race			
African American	1 = African American	55%	
Caucasian	2 = Caucasian	32%	
Others	3 = Others	13%	
Parent's education level		7.91	2.03
Less than high school	1 = Grade 9 to 10 = BA	8%	
High school		22%	
More than high school		70%	
Control variables			
English as a second language	1 = Yes; 0 = No	7%	
Number in the family	1 = Less than 5; 0 = More than 5	74%	
Site level characteristics			
Star rating	Range 1 to 3	2.38	.82

Note. The standard deviation is not reported for binary variables (i.e., female, race, and number in the family). PPVT = Peabody Picture Vocabulary Test. WJ-III = Woodcock Johnson-III. LW = Letter/word recognition. AP = Applied problems.

TABLE 3
Fixed and Random Effect Estimates for Universal Prekindergarten Unconditional Model

	PPVT-IV		*WJ-III-LW*		*WJ-III-AP*	
Fixed Effect	*b*	*SE*	*b*	*SE*	*b*	*SE*
Average initial status	97.48***	1.42	102.19***	1.12	101.62***	1.42
Average change in test score	.92*	.37	.86*	.39	.70	.40
BIC	4023.4		3956.8		4134.4	
Random Effect	Variance Component		Variance Component		Variance Component	
Level-1 variance						
Residual variance	28.62***		29.60***		35.10***	
Level-2 (Children within site)						
Children initial status	184.57***		138.50***		130.53***	
Average change in score	10.77***		12.84***		11.88***	
Level-3 (Between sites)						
School average score	24.06*		12.64*		17.08*	

Note. PPVT = Peabody Picture Vocabulary Test. WJ-III = Woodcock Johnson-III. LW = Letter/word recognition. AP = Applied problems. *SE* = standard error. BIC = Bayesian Information Criteria.
*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

time for the PPVT-IV and WJ-III-LW. The random effect estimates at level 2 indicate significant variability among individuals with respect to their initial test score and also in terms of average change in their test scores. The random effect estimates for level 3 also suggest statistically significant variability among sites in terms of the initial average test score. However, the variability in the average change in the test score among sites is not statistically significant (not reported in the table). In addition, the inclusion of random variation at level 3 did not increase the model fit. Hence, the random variation term for average change was not included at level 3.

Results by Achievement Test

The results of the model were examined using each of the three achievement test scores at the third time point as the dependent variable.

PPVT-IV. As seen in Table 4, regression coefficients indicate that African American children and children from other races begin with significantly lower test scores than White children on the PPVT-IV. However, African American children tend to perform better over time, compared to White children, showing gains of 1.77 more points over time compared to those of White children. Similarly, children whose parents' education level is higher tend to start off significantly better and to show greater gains than their counterparts.

Comparisons of the random effects from the unconditional model to the full model in Table 5 indicate that there is significant change after the inclusion of the level-2 variables for the PPVT-IV. Almost 40% of the variation in the initial test score is explained by level-2 variables ([184.57 – 111.53]/184.57), whereas the level-2 variables explained only 13% of the variation in the average change among children within site ([10.77 – 9.33]/10.77). The star rating does not explain a significant amount of variation in average PPVT-IV score among sites.

TABLE 4
Fixed Effect Estimates and Final Model Predicting Achievement Scores

	PPVT-IV Estimates		*WJ-LW Estimates*		*WJ-AP Estimates*	
Fixed Effect	*b*	*SE*	*b*	*SE*	*b*	*SE*
Model for initial status						
Model for average score within site						
Intercept	104.73***	2.41	99.84***	2.56	106.97***	2.55
Star rating	.60	1.62	.88	2.04	1.12	1.98
Model for variables on initial status						
Gender	.20	1.67	2.47	1.54	1.76	1.60
African American	–14.19***	1.97	–3.44	2.09	–10.13***	2.09
Other race	–10.40**	3.69	–1.67	3.49	–6.43	3.46
Education	.89*	.45	.75	.42	1.03*	.43
English as second language	–10.19**	4.91	–12.44**	4.56	–8.93	4.63
Number in family	3.09	1.91	4.59**	1.76	.29	1.83
Model for average change in score within site						
Intercept	–.96	.92	.16	1.01	–1.53	1.00
Gender	.27	.73	–.70	.79	–.06	.80
African American	1.77*	.79	.28	.87	1.40	.87
Other race	2.31	1.67	1.68	1.86	5.82**	1.77
Education	.55*	.21	–.13	.23	–.03	.23
English as second language	.11	2.21	–.59	2.41	–6.39**	2.37
Number in family	.41	.85	1.02	.93	1.74	.93
BIC	3769.60		3755.70		3920.4	

Note. PPVT = Peabody Picture Vocabulary Test. WJ-III = Woodcock Johnson-III. LW = Letter/word recognition. AP = Applied problems. SE = Standard Error; BIC = Bayesian Information Criteria.
*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

WJ-III-LW. As seen in Table 4, regression estimates for average initial WJ-III-LW score within site suggest that children whose native language is not English score, on average, 12 points lower than those whose native language is English. Children from smaller families showed

TABLE 5
Three-Level Random Effect Estimates of Achievement Tests

Random Effect	*PPVBT-IV Estimates Variance Component*	*WJ-LW Estimates Variance Component*	*WJ-AP Estimates Variance Component*
Level-1 variance			
Residual variance	27.39***	30.04***	34.86***
Level-2 (Children within site)			
Initial status	111.53**	84.83***	90.81***
Average change in score	9.33**	11.88***	10.29**
Level-3 (Between sites)			
School average score	.69	9.55	6.51

Note. PPVT = Peabody Picture Vocabulary Test. WJ-III = Woodcock Johnson-III. LW = Letter/word recognition. AP = Applied problems.
*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

higher scores at baseline. However, none of these predictors has significant influence on the average change in test scores within site. Examination of the random effects model compared to the full model in Table 5 shows that almost 39% of variation in initial average test scores within site is explained by predictors included in the full model. The same set of predictors only explains 7% of variation in the average change in test scores within each site. The star rating does not explain significant variation in average WJ-III-LW score among sites.

WJ-III-AP. As seen in Table 4, the regression estimate for African American children shows that they begin the study with initial average scores 10.13 points lower than White children. Also, children whose parents have higher education levels scored 1.03 points higher at baseline than their counterparts. The examination of this same set of predictors on average change in test score within site reveals that children who are not of African American origin tend to perform far better than White children, as reflected by an average 5.82-point increase in their test scores over time. However, children whose native language is not English tend to perform poorly over time on this test, compared to English speakers, as reflected by the negative magnitude of the regression estimate. The change evident between the random effects model and the full model in Table 5 illustrate that almost 40% of the variation in the initial average test score within site is explained by the predictors included in the full model. However, the same list of variables explains only 13% of the variation in the average change in the test score within site. The star rating does not explain significant variation in average WJ-III-AP score among sites.

DISCUSSION

In this study, we examined results of a UPK pilot study to determine if students' achievement scores on three different achievement measures improved over time, and what factors predicted change in scores over time. Results indicate that achievement scores of participating children on two of the three measures (i.e., PPVT-IV and WJ-III-LW) improved significantly over the course of the study. Changes in achievement scores were shown to be significantly predicted by race, parental education level, and whether the family spoke English as a second language. Contrary to expectations, the quality of the prekindergarten setting failed to predict change over time in achievement scores.

The results suggest that achievement improves more for children who were most at risk at baseline (i.e., who started below the 50th percentile on the achievement test). Further, the magnitude of the gains for the most at-risk children exceeds the gains to be expected from having any preschool experience, and are comparable to the gains found in larger-scale initiatives focused on providing high quality UPK.

The lack of significant findings related to the quality ratings (i.e., star rating) could be due to the rating scale not being sensitive enough to reflect differences between sites with higher and lower structural quality. Relatedly, because the pilot sought to raise quality in these sites, the sample reflected a fairly narrow range of quality. Future investigation into UPK would benefit from inclusion of stronger rating scales that are both sensitive to differences and comprehensive in the range of quality assessed.

That no significant change was found for achievement over time using the WJ-III-AP mirrors what has been found in past studies. Of four experimental or quasi-experimental studies reviewed

by Brown and Scott-Little (2003), only one reported significant results for math and logic skills. In contrast, of the four experimental or quasi-experimental studies, three of them reported significant results for language and literacy skills. This suggests that children are better able to improve in language and literacy skills than in math and logic skills. As a consequence of this trend, it may be important for educators responsible for curriculum development to reassess the curriculum content in regard to math and logic skills.

A weakness in the study design used here is the lack of a control group, as well as the lack of being able to control for several variables across the different sites. As a consequence of these limitations, conclusions regarding the effect of prekindergarten on children's achievement scores should be made cautiously. Moreover, because of significant variation between sites, one cannot determine if other variables may have influenced the outcome. For example, because sites were able to use different curricula, the curriculum being used may have affected improvement on achievement scores. Finally, despite the strong psychometric properties of the PPVT-III, findings from a recent study suggest that African American children may score lower on this measure than European American children due to unintentional bias in the measure (Restrepo et al., 2006). This is an important consideration, given the large percentage of African American children in the sample for this study.

Despite study limitations, this pilot project contributes to the growing body of knowledge that children who begin in lower achievement brackets make the strongest gains over time. Particularly important for educators and policy makers to understand is whether or not achievement scores are the best measure of success for preschool generally and UPK projects specifically. The often-cited Perry Preschool Project followed a sample of children longitudinally for 3 decades, and identified factors such as health (Muennig, Schweinhart, Montie, & Neidell, 2009) and economic benefits (Schweinhart, Barnes, & Weikart, 1993) for children attending preschool. Muennig et al. (2009) found that several positive characteristics associated with these children, including educational attainment, earnings, and stable family environments, predicted improvements in their health behaviors when they reached age 40 compared to children who were not part of the study. Schweinhart and Weikart (1998) summarized the economic benefits of Perry Preschool:

> A high-quality preschool program cuts participants' life-time arrest rate in half, significantly improves their educational and subsequent economic success, and provides taxpayers a return equal to 716 percent of their original investment in the program, a return that outperformed the US stock market during the same period of time. (p. 57)

Although the Perry Preschool Project was a tightly controlled study that would be difficult to duplicate on a large scale, it nonetheless illustrates that the most important benefits may not, in fact, be improved achievement scores, but rather an overall improved ability to contribute to society. Consequently, it may be important to consider measures that assess other skills (e.g., motivation) that may result from preschool education.

In addition to the contributions to the broader knowledge base, this study has served to inform the planning for UPK in the pilot county. In general, the findings supported the continuation of the pilot with some refinements. One key change was to adopt a developmental measure that could provide short-term feedback to prekindergarten teachers about the emerging abilities of their students. The program selected the Bracken School Readiness Assessment (Panter & Bracken, 2009), which provides immediate feedback to teachers and parents about the child's abilities and recommends customized strategies for enhancing these abilities. The program has

also worked with its participating sites to clarify how selected curricula are used, and how sites are pursuing quality enhancement activities to improve the learning environment for children. The pilot has completed 5 years of implementation, serving approximately 1,000 children per year, and has recently been expanded to a capacity of 1,500 children. The study of the program is following children forward to kindergarten to examine children's performance on the state-mandated school readiness assessment (i.e., Kindergarten Readiness Assessment–Literacy) and how this correlates with their experience in a UPK setting.

ACKNOWLEDGMENTS

This study was funded through the Office of Early Childhood/Invest in Children of Cuyahoga County, Ohio, with support from The Cleveland Foundation. We thank all the families who took part in the study. We also thank Dr. Rebekah Dorman and Robert Staib of the Office of Early Childhood for key support and input on the study.

REFERENCES

Abbott-Shim, M., Lambert, R., & McCarty, F. (2003). A comparison of school readiness outcomes for children randomly assigned to a Head Start program and the program's wait list. *Journal of Education for Students Placed At Risk*, *8*, 191–214. Retrieved from http://www.tandfonline.com/toc/hjsp20/current

Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *Future of Children*, *5*(3), 25–50. Retrieved from http://www.princeton.edu/futureofchildren/index.xml

Barnett, W. S., & Lamy, C. E. (2006). *Estimated impacts of number of years of preschool attendance on vocabulary, literacy and math skills at kindergarten entry* (Working Paper). Brunswick, NJ: National Institute for Early Education Research. Retrieved from http://nieer.org/

Brown, E. G., & Scott-Little, C. (2003, March). *Evaluations of school readiness initiatives: What are we learning? SERVE's expanded learning opportunities national leadership area research report.* Greensboro, NC: SouthEastern Regional Vision for Education.

Bryant, D., Maxwell, K., Taylor, D., Poe, M., Peisner-Feinberg, E., & Bernier, K. (2003). *Smart Start and preschool child care quality in North Carolina: Change over time and relation to children's readiness.* Chapel Hill, NC: Frank Porter Graham Child Development Institute. Retrieved from http://fpg.unc.edu/node/4610

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test Manual* (4th ed). Minneapolis, MN: Pearson.

Gormley, W. T., & Phillips, D. (2005). The effects of universal pre-K in Oklahoma: Research highlights and policy implications. *Policy Studies Journal*, *33*(1), 65–82. doi: 10.1111/j.1541-0072.2005.00092.x.

Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early Childhood Environment Rating Scale: Revised edition*. New York, NY: Teachers College Press.

Henry, G. T., Henderson, L. W., Ponder, B. D., Gordon, C. S., Mashburn, A. J., & Rickman, D. K. (2003). *Report of the findings from the Early Childhood Study: 2001–2002.* Retrieved from http://aysps.gsu.edu/publications/2003/earlychildhood.htm

Karoly, L. A., Greenwood, P. W., Everingham, S. S., Hoube, J., Kilburn, M. R., Rydell, C. P., … Chiesa, J. (1998). *Investing in our children: What we know and don't know about the costs and benefits of early childhood interventions.* Santa Monica, CA: Rand.

Kirp, D. (2005). Before school. *Nation*, *28*(17), 24–30.

McCall, R. B., Larsen, L., & Ingram, A. (2003). The science and policies of early childhood education and family services. In A. J. Reynolds, M. C. Wang, & H. J. Walberg (Eds.), *Early childhood programs for a new century* (pp. 255–298). Washington, DC: Child Welfare League of America Press.

Muennig, P., Schweinhart, L., Montie, J., & Neidell, M. (2009). Effects of a prekindergarten educational intervention on adult health: 37-year follow-up results of a randomized controlled trial. *American Journal of Public Health*, *99*, 1431–1437. Retrieved from http://ajph.aphapublications.org/

Olsen, L., & DeBoise, T. (2007). Enhancing school readiness: The Early Head Start model. *Children & School*, *29*(1), 47–50. Retrieved from http://cs.oxfordjournals.org/

Panter, J. E., & Bracken, B. A. (2009). Validity of the Bracken School Readiness Assessment for predicting first grade readiness. *Psychology in the Schools*, *46*, 397–409. doi: 10.1002/pits.20385

Restrepo, M. A., Schwanenflugel, P. J., Blake, J., Neuharth-Pritchett, S., Cramer, S. E., & Ruston, H. P. (2006). Performance on the PPVT-III and the EVT: Applicability of the measures with African American and European American preschool children. *Language*, *Speech*, *and Hearing Services in Schools*, *37*, 17–27. doi: 10.1044/0161-1461(2006/003).

Schumacher, R., Ewen, D., & Hart, K. (2005). *All together now: State experiences in using community-based child care to provide pre-kindergarten.* Washington, DC: Center for Law and Social Policy.

Schweinhart, L. J., Barnes, H. V., & Weikart, D. P. (1993). *Significant benefits: The High/Scope Perry Preschool Study through age 27* (Monographs of the High/Scope Educational Research Foundation, 10). Ypsilanti, MI: High/Scope Press. Retrieved from http://www.highscope.org/

Schweinhart, L. J., & Weikart, D. P. (1998). Why curriculum matters in early childhood education. *Educational Leadership*, *55*(6), 57–60. Retrieved from http://www.ascd.org/publications/educational-leadership.aspx

Woodcock, R. W., McGrew, K. S., Schrank, F. A., & Mather, N. (2001, 2007). *Woodcock-Johnson III Normative Update.* Rolling Meadows, IL: Riverside.

Zill, N., Resnick, G., Kim, K., O'Donnell, K., & Sorongon, A. (2003). *Head Start FACES 2000: A whole-child perspective on program performance.* Retrieved from http://www.acf.hhs.gov/.