Abbott Preschool Program Longitudinal Effects Study: Fifth Grade Follow-Up

W. Steven Barnett, Ph.D.
Kwanghee Jung, Ph.D.
Min-Jong Youn, Ph.D.
Ellen C. Frede, Ph.D. (Acelero, Inc.)

March 20, 2013

National Institute for Early Education Research
Rutgers—The State University of New Jersey
www.nieer.org

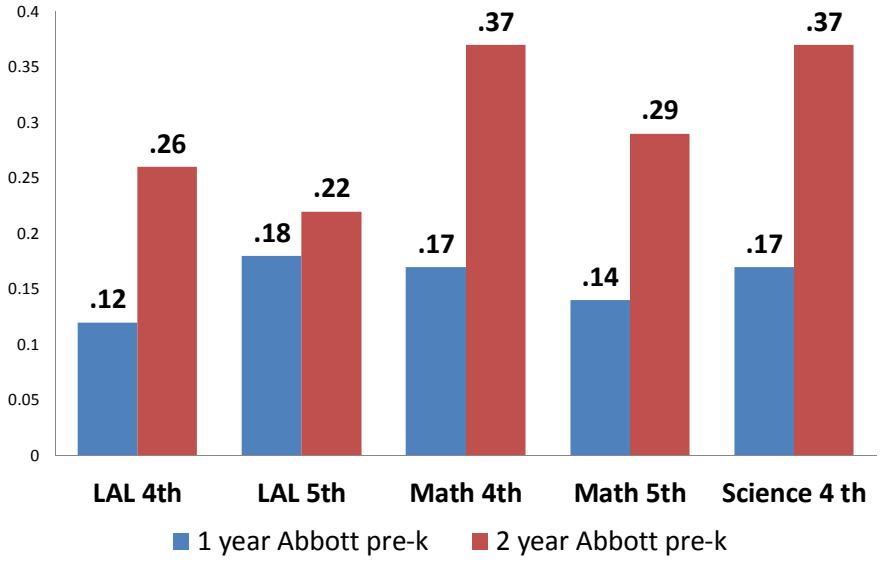
Abbott Preschool Program Longitudinal Effects Study (APPLES): Fifth Grade Follow-Up Executive Summary

New Jersey's Abbott Preschool program is of broad national and international interest because the Abbott program provides a model for building a high-quality system of universal pre-K through public-private partnerships that transform the existing system. The program offers high-quality pre-K to all children in 31 New Jersey communities with high levels of poverty and about a quarter of the state's children. The Abbott Preschool Program Longitudinal Effects Study (APPLES) assesses the impact of this pre-K program on children's learning and development based on a cohort of children who completed their 4-year-old year in 2004-05.

APPLES previously estimated the impacts of Abbott pre-K at kindergarten entry and second grade follow-up. We found substantial impacts on individually administered assessments of language, literacy, and mathematics at both times. In addition, pre-K was found to reduce grade retention. Previous analysis also indicated that APPLES methodology tends to underestimate impacts. Moreover, as pre-K quality continued to rise after 2004-05, even perfect estimates would underestimate the effect of the program in more recent years.

The 4th and 5th grade APPLES follow-up finds that Abbott preschool programs increased achievement in Language Arts and Literacy, Math, and Science, as shown in Figure 1 below. Our estimates indicate that two-years of pre-K beginning at age 3 had larger persistent effects on achievement than did one year of pre-K. The magnitude of the test score gains from one year are equivalent to roughly 10 to 20 percent of the achievement gap between minority and white students. The gains from two years are equivalent to 20 to 40 percent of the achievement gap. These gains are an even larger portion of the typical learning gain that occurs in year.

Figure 1. Pre-K Abbott Effects on NJASK by Years of Participation

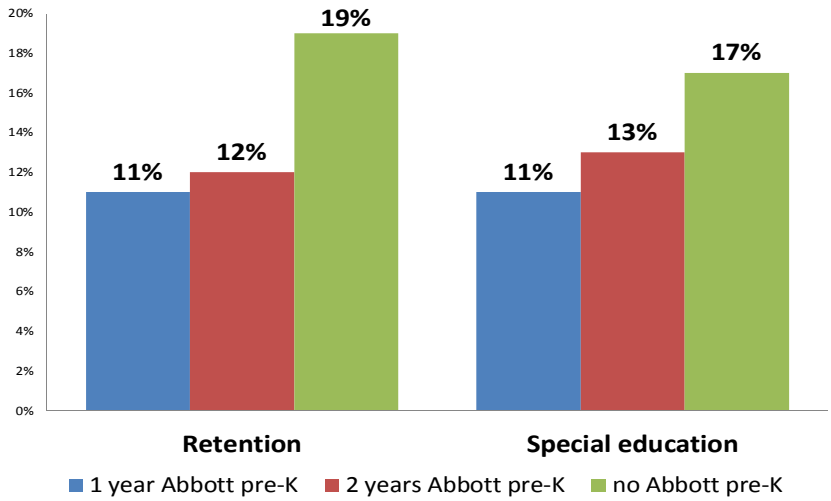


The Abbott Preschool program also is found to decrease grade retention and special education placement rates. However, we do not find that two years of pre-K leads to larger gains than does one year. These results are displayed in Figure 2 below (percentages are adjusted to equalize for differences in child background characteristics). Decreases in grade repetition and special education can significantly reduce the costs of education from kindergarten through 12th grade, and a substantial percentage of the cost of high-quality pre-K may be recovered as a result.

The Abbott Preschool program's effects on achievement and school success are larger than has been found for less well-funded programs with weaker standards. The gains in achievement from two years of the program are similar in size to those of the Chicago Child Parent Centers, as are the impacts of the Abbott pre-K program overall on grade retention and special education. It is noteworthy that the Abbott program serves all children in the Abbott districts, not just those who are low income. The Chicago program served a somewhat higher percentage of low-income families. Also, the gains found are particularly impressive as the Abbott program had not yet reached its ultimate level of quality (as documented in a series of studies).

Based on the results of the 5th grade follow-up we suggest that New Jersey would be wise to take additional steps to build on its success to date. First, participation rates in two-years of pre-K should be increased to above 90 percent in the Abbott districts. In some districts it is far below this level and the human costs in low achievement and school failure is tragic as is the likely adult consequences of lower productivity and earnings and higher crime rates. Second, as required by the New Jersey School Funding Reform Act of 2008, high-quality pre-K should be expanded to offer a comparable program to all low-income children. In addition, plans should be developed to extend the opportunity for high-quality pre-K to all of the state's children.

Figure 2. Pre-K Abbott Effects on Retention and Special Education



Abbott Preschool Program Longitudinal Effects Study (APPLES): Fifth Grade Follow-Up

Introduction

This study investigates the persistence of educational effects of state-funded pre-K for children at ages 3 and 4 in New Jersey's Abbott districts through 5th grade. The pre-K program was developed in response to the landmark New Jersey Supreme Court school-funding case, *Abbott v. Burke*. In the fall of 1999, 3- and 4-year-old children in 31 high poverty districts began to enroll in this new high-quality preschool education program. This program was designed to prepare children to enter school with the knowledge and skills necessary to meet the New Jersey Preschool Teaching and Learning Expectations: Standards of Quality (NJ DOE, 2002b) and the Kindergarten New Jersey Core Curriculum Content Standards (NJ DOE, 2002a). Through a Department of Education (DOE) and Department of Human Services (DHS) partnership, Abbott preschool classrooms combined a DOE-funded six-hour, 180-day component with a DHS-funded wrap-around program that provided daily before- and after-care and summer programs. In total, the full-day, full-year program was available up to 10 hours per day, 245 days a year.

Enrollment in the Abbott preschool program increased rapidly over the next five years followed by more modest growth. In the 31 former Abbott districts more than 43,000 3- and 4-year-old children (about 80 percent of the population in those districts) have been served annually in recent years. Pre-K is delivered by a mixed public-private delivery system overseen by the public schools. Private centers and Head Start agencies contract with local boards of education to serve about two-thirds of the children. The rest are served in public school classrooms.

Basic program standards for this pre-K program include a maximum class size of 15, teachers with certification in early childhood education, assistant teachers in every classroom,

support services for children and families, and a developmentally appropriate curriculum that fully addresses the state's learning standards. Some of these standards could be implemented quite quickly. Others, like the requirement for teacher certification, took time. Over the first five years, everyone from the classroom level on up worked hard to fully implement these standards and to bring classroom practices up to the level of effectiveness that these standards were designed to support (Frede & Barnett, 2011). Such standards facilitate highly effective preschool education, but do not by themselves guarantee it. To ensure high quality and consistency for children across auspice and district and to assist administrators and staff who may have been inadequately prepared in early childhood education, more detailed operational standards were developed (Abbott Preschool Program Implementation Guidelines; Office of Early Childhood Education, NJ DOE, 2002, revised 2005). These standards were also designed to ensure that the particular needs of children in each community were addressed. The Abbott preschool program was not designed to be simply a "cookie cutter" approach that is identical in every community.

Rather than replacing the existing system, New Jersey's approach was to transform the existing system, providing education and training to improve teachers' expertise as well as boosting pay and raising standards. Observation data on preschool classroom quality were collected on pre-K in the Abbott districts beginning in the spring of 2000. By the 2004-2005 cohort studied here quality had improved sufficiently that it was judged useful to assess the impacts of the pre-K program on children's learning (Lamy, Frede, & ELIC, 2005). On the Early Childhood Environment Rating Scale-Revised (ECERS-R; Harms, Clifford, & Cryer, 1998) pre-K classroom quality rose from an average of 3.86 in 2000, with 24 percent scoring below 3 (minimal), to 4.77 in spring 2005 with only about 2 percent scoring below 3. The

system transformation was not complete by 2004-2005 as 22 percent of teachers still lacked full certification, and quality continued to improve with the average ECERS-R score reaching 5.2 by 2008 and 63 percent of pre-K classrooms scoring above 5, in the good to excellent range (Frede, Jung, Barnett, & Figueras, 2009). This strongly indicates that our study of the 2004-2005 cohort is likely to underestimate the impacts on achievement and school success of the Abbott pre-K model as it is fully implemented today.

Review of Research on Preschool with Emphasis on Public Pre-K

Over the past 50 years, much research has accumulated regarding the effects of preschool education. Far too often this research is selectively cited by proponents of one view or another to confirm their pre-existing views. A more responsible approach is to scrutinize all of the evidence regardless of its findings. Meta-analysis provides a straightforward statistical summary of findings across all studies and identifies reasons that results vary across studies. A recent and comprehensive meta-analysis of the immediate and long-term effects of preschool education is provided by Camilli, Vargas, Ryan, and Barnett (2010). Consistent with previous summaries, this meta-analysis finds that preschool education has substantial effects. Effects diminish, but do not disappear, after children leave the program and enter primary school. Cognitive effects persist throughout the school years at about half the size of the initial effects. No significant differences over time are found for other types of effects, but fewer studies measure social and emotional outcomes while other outcomes such as grade repetition or high school graduation rates are only measured in the long term.

Camilli et al. (2010) provide some insights into why some studies find larger cognitive gains than do others. More rigorous studies, including randomized trials, find larger impacts. Programs emphasizing intentional teaching and individualization (small groups and one-to-one

interaction) also are associated with larger cognitive effects. Unexpectedly, providing added social and health services was associated with smaller impacts, perhaps because program resources are stretched too thinly, perhaps because such services are available elsewhere.

The recent national impact study has brought increased attention to the notion of "fadeout" in effects (U.S. Department of Health and Human Services, 2010). This study has been incorrectly cited as evidence that large scale preschool programs do not produce persistent effects. Most attention has focused on this study's "intent-to-treat" estimates. Such comparisons estimate the effects of whether a child was randomly assigned to Head Start or not rather than the effects of Head Start *per se*. Some children assigned to Head Start did not attend; some assigned to the control group did attend. In addition, some control group members attended other preschool programs. The result is that the intent-to-treat estimates are likely to underestimate Head Start's actual effects by more than 50 percent. In addition, it must be recognized that control children can catch up with Head Start attendees later on because of the compensatory efforts of public schools. This "catch up" effect was observed in other randomized trials, including the well-known Perry Preschool study, and should not be confused with "fade-out" (Barnett, 2011).

To put the Abbott pre-K study in context, our review focuses primarily on studies of preschool programs operating under the auspices of public education at the state or local level. This is a fairly diverse group of programs as they differ considerably in the populations served, standards, and funding levels from one state and, sometimes, locality to another (Barnett, et al. 2010). State pre-K results are very likely just as diverse, and to date there has been no evaluation of a representative sample of such programs. However, there have been more than a few studies and their results are worth reviewing as context for the present study.

The oldest studies are of local programs, sometimes funded as part of federal initiatives. These include the Perry Preschool and Institute for Developmental Studies programs that were operated by the public schools (Consortium for Longitudinal Studies, 1983; Deutsch, Deutsch, Jordan, & Grallow, 1983; Deutsch, Taleporos, & Victor, 1974; Jordan, Grallo, Deutsch, & Deutsch, 1985; Schweinhart et al., 2005). More recently the Chicago Longitudinal study has provided highly relevant evidence because it is a rigorous long-term study of a large-scale program, the Child Parent Centers (CPC), operated by the Chicago public schools over many years (Temple & Reynolds, 2007).

The CPCs were similar in design and cost to many of today's stronger state-funded preschool programs with a certified teacher and an assistant in each classroom of 18 children, and a parent outreach and support component. Estimated CPC effects on test scores at kindergarten entry were .35-.77SD for all children and .20-.65SD for those who attended only one year (Reynolds, 2000). Effects on cognitive abilities subsequently remained about .20SD through eighth grade. (To permit comparisons across studies, all effects are reported as effect sizes in standard deviation (SD) units. As the achievement gap for minority children is about 1SD, these effects can be interpreted as a percentage of the achievement gap.) Effects on other outcomes include significant reductions in grade retention, special education placements, and both juvenile and adult arrests, and a significant increase in high school graduation.

More recent studies provide rigorous estimates of the effects of today's large scale, state-funded pre-K (Barnett, Howes, & Jung, 2008; Gormley, Gayer, Phillips, & Dawson, 2005; Gormley, Phillips & Gayer, 2008; Hustedt, Barnett, Jung, & Figueras, 2008; Hustedt, Barnett, Jung, & Thomas, 2007; Lipsey, Farran, Hofer, Bilbrey & Dong, 2011; Wong, Cook, Barnett & Jung, 2008). These studies find positive effects regardless of the family incomes of the children

served, though effects tend to be somewhat larger for children from lower-income families. Most prior research in the United States focused on programs for disadvantaged children, so new state studies add significantly to our knowledge about impacts on more general populations. However, there are exceptions even in the older research including the BYU randomized trial of pre-K for relatively advantaged children that found large cognitive gains immediately and into the early school years (Larsen & Robinson, 1989).

Many recent studies of state funded pre-K have used a regression discontinuity design (RDD) to estimate impacts, particularly for programs designed to serve all children. Some have claimed that this design produces estimates that are upward biased (Whitehurst, 2013). No evidence is offered to support these claims, which seem to arise from a concern that they produce higher estimates of impacts than the national randomized trial of Head Start. Yet, there is evidence that readily explains this discrepancy and validates the RDD design. First, as already noted, the National Impact Study's primary estimates are biased downward. The RDD design avoids the problems of crossovers and participation in other programs. Second, the state pre-K programs studied may well be more effective in producing cognitive gains as they tend to have better paid, more qualified teachers (as discussed below). A study of Tennessee's state pre-K program that employed a randomized trial and an RDD finds relatively large and similar impacts using both methodologies, but the RDD produces somewhat larger estimates (Lipsey et al., 2011).

Rigorous long-term follow-up studies of state pre-K programs also have been conducted, but most continue only a few years into the primary grades. One of the earliest such studies found positive effects of New York's program on general reasoning, verbal concepts, and school-related knowledge and skills as well as retention in grade (Irvine, Horan, Flint, Kukuk, &

Hick, 1982). Other states with rigorous evidence that impacts persist well into the school years include Georgia, Oklahoma, Michigan, North Carolina, South Carolina, and Texas (Fitzpatrick, 2008; Frede & Barnett, 1992; Gormley et al., 2008; Kuhne, 2008; Malofeeva, Daniel-Echol & Xiang, 2007; Peisner-Feinberg & Schaaf, 2010). These studies differ in the size of the impacts they find as well as in length of follow-up and methodological strengths and weaknesses. However, all find some persistent effect on achievement.

The longest recent follow-up of state pre-K is provided by a study of Michigan's Great Start School Readiness program. This study finds pre-K associated with test score gains through 4th grade and again on the high school exit exam in 11th and 12th grades, though not on the state's middle school exams (Schweinhart, Xiang, Daniel-Echols, Browning, & Wakabayashi, 2012). This Michigan study also finds that state pre-K led to less grade retention through 12th grade and to more on-time high school graduation. The estimated reduction in grade retention generated cost-savings equal to about 40 percent of the cost of the pre-K program.

Study Design and Earlier APPLES Findings

In the fall of 2005, we implemented a two-step research process to estimate the long-term effects of attendance in an Abbott preschool classroom. The first step was to implement a Regression Discontinuity Design (RDD) to estimate the effects of the program on children's abilities at kindergarten entry (Trochim, 1984). This approach relies on the fact that eligibility for Abbott pre-K within a designated school district is determined by date of birth alone. This assignment rule allows us to construct two groups, one entering kindergarten that has already attended the program at age 4 and one entering preschool that has not yet attended at age 4. These groups are unlikely to differ with respect to measured or unmeasured child and family characteristics so that the RDD minimizes the potential effects of selection bias, which occurs

when the effects of differences between the two groups of children are confounded with program effects (Cook, 2008; Wong, Cook, Barnett, & Jung, 2008).

Unfortunately, the RDD approach cannot provide an estimate of effects beyond kindergarten entry. We employed a second design to obtain estimates beyond kindergarten entry – comparing children who attended pre-K to a conventional comparison group who did not attend the Abbott pre-K program and identified at kindergarten entry. We then assessed the accuracy of estimates obtained from this second approach at kindergarten entry by comparing them to the RDD estimates. If the initial estimates from both analyses are similar, then we can have confidence in the longitudinal results. To the extent that they differ, we have an indication of the likely direction and magnitude of bias in the longitudinal estimates. One caveat is that in the RDD design both groups have the same prior experiences and attend the same pre-K program so that the comparison group can be viewed as a true "no-preschool" group whereas in the longitudinal comparison group some children attended other preschool programs at ages 3 and 4. For this reason, the longitudinal study might be expected to produce somewhat smaller estimates of effects on learning and development.

Our sample for the second design used for APPLES consists of children from the RDD study who attended the Abbott Preschool program and an additional comparison sample of children who attended the same kindergarten classrooms as the pre-K children, but who did not attend the Abbott preschool program. The sample was drawn proportionately from the 15 largest Abbott districts in order to minimize data collection costs and based on the knowledge that the smaller districts had somewhat higher quality (Frede, Jung, Barnett, Lamy, & Figueras, 2007). Not only does our sample represent the vast majority of children who participated in the program, but modestly underrepresents quality across all of the districts. We obtained a sample

at kindergarten entry of 1,038 children in 15 districts. Of these, 284 did not attend the Abbott pre-K program, 451 attended for one year, and 303 attended for two years. As some children attended Abbott preschool for one year at age 4 and others attended preschool for two years at ages 3 and 4, we are able to separately estimate the effects of one year and two years of preschool attendance using this second design. Treatment and comparison samples are balanced with respect to school district, and we control for district in the analyses.

Both approaches find positive effects on children's learning in the areas of oral language, early literacy, and mathematics at kindergarten entry. We present standardized effects, which are converted to standard deviation (SD) units for comparison over time and across measures. The estimated effects using the RDD design for one year of pre-K at age 4 were .28SD for language, .56SD for print awareness, and .36SD for math. The estimated effects for one year of pre-K using the longitudinal study's (APPLES) conventional comparison group were .21SD for language, .29SD for print awareness, and .20SD for math. Comparing these RDD and APPLES estimates for one year of pre-K at kindergarten entry indicates that the APPLES design underestimates the effects of the Abbott program. Standardized effect sizes from the APPLES longitudinal study design at kindergarten entry were larger for two years of participation than for one year, .42SD for language, .31SD for print awareness, and .34SD for math. The RDD design cannot provide estimates for two years of pre-K participation.

Subsequent follow-ups of the longitudinal sample were conducted in the spring of 2007 and 2008. Children would have been in first and second grade, respectively, if they had not been retained in grade at any point. Children were assessed with their age cohort regardless of their actual grade level, and the 2007 data are referred to as "first grade" and the 2008 data as "second grade" even though some of the children in each year's data were actually behind a grade level.

In addition, we collected demographic data on the sample including age, gender, ethnicity, and lunch status. Although some children could not be found in the follow-up, attrition analyses revealed no significant differences between the initial sample and the 2nd grade follow-up sample with respect to demographics or scores at kindergarten entry.

At the end of 2nd grade the effects of Abbott Pre-K participation continued to be significant with effects of .22SD ($p<.05$) for one year of attendance and .40SD ($p<.01$) for two years on receptive vocabulary. Effects on mathematics at the end of 2nd grade were .24SD ($p<.05$) for one year and .44SD ($p<.01$) for two years. At the end of 2nd grade, effects on literacy development were most apparent on Passage Comprehension, on which the former pre-K attendees scored higher with effect sizes equal to .16SD for one year and .20SD for two years. In addition, effects were found for grade retention, which was 10.7 percent for children who did not attend pre-K, 7.2 percent for those who attended for one year, and 5.3 percent for those who attended two years. For additional details regarding measures and findings at 2nd grade, readers are referred to our earlier APPLES Blossom report (Frede et al., 2009).

Fifth Grade Follow-up Data and Analysis

To estimate effects beyond grade 2, we relied upon state of New Jersey standardized tests that are administered statewide beginning in grade 3. We were able to successfully exactly match 754 (70%) children from the original sample against the state database for children who were tested in grades 3 to 5. The follow-up sample includes 553 (72%) of the original 766 children who attended Abbott Preschool and 201 (66%) of the original 305 who did not attend. Attrition analyses indicate that the difference in follow-up sample size could have occurred by chance and finds that the students for whom we have follow-up test scores are highly similar in

their demographic characteristics and scores at kindergarten entry to those not matched against the state assessment database.

Test score data comes from the New Jersey Assessment of Skills and Knowledge (NJASK) for three years: 2009 (3rd grade), 2010 (4th grade), and 2011 (5th grade). The state of New Jersey graciously provided data for our sample to make this follow-up possible. The state NJASK database allows us to include scores for children who had been retained in-grade for analysis with their age cohort as if they had not been retained. This provided a larger and more robust follow-up sample and prevented the introduction of bias due to grade retention that would systematically omit scores for children who lagged most behind. We focus on the 4th and 5th grade assessments as we were able to account for virtually all repeaters in the 5th grade analyses and most repeaters in the 4th grade analyses.

Measures

Child and Family Characteristics. Demographic and family background characteristics available for all children from the assessment database include age, gender, ethnicity, and free lunch status at the time of the statewide assessment. In addition, a parent survey at kindergarten provided for 512 of the 754 participants parental education, marital status of the child's mother or guardian, parental employment status, and home language. These demographic data are reported for each analytical subsample in Table 1. Descriptive statistics for kindergarten year data are reported in Table 2 for each group by whether or not they were found in the 3rd to 5th grade follow-up.

For analysis we coded race/ethnicity as Black, Hispanic, White, and Other; White was the reference group to which the others were compared. Gender of child was dummy coded with male as the reference group. Age was measured in months at 3rd grade. Parental employment

status was dummy coded with the reference group those not working. Parental education was coded: (1) less than high school, (2) high school degree or GED, (3) some college or post-high school technical school, and (4) BA degree or better. Family structure was dummy coded as married or single parent/guardian with single parent/guardian the reference group. Language was dummy coded with non-English as the reference group.

Educational Achievement and School Success. The NJASK assesses achievement in three domains: language arts and literacy, math, and science. The Language Arts Literacy (LAL) test measures a student's reading comprehension and writing abilities. It comprises reading passages, multiple-choice items, constructed-response items, and writing tasks. The Mathematics test measures students' ability to solve problems by applying mathematical concepts. The Science test measures 4th grade students' abilities to recall information and solve problems by applying science concepts. The Science test comprises multiple-choice and constructed response items and assesses knowledge and application skills in Life Science, Earth Science, and Physical Science. The Science test was administered only at 4th grade.

The state's assessment database also allowed us to classify children as having experienced retention (behind expected grade based on the year and grade for which a test score is reported) or having a special education placement in grades 3 through 5. To include children who were retained in-grade (and did not have later NJASK scores in the database) in their cohort's analysis of 4th and 5th grade scores we "predicted" scores for those children. Predicted 4th and 5th grade scores were estimated from regressions of 4th and 5th grade scores on prior year scores controlling for child and family characteristics. This may be regarded as rescaling the earlier grade score to its equivalent in the subsequent grade at the same age.

Data Analysis

The participants in this study are clustered (in groups) by virtue of their belonging to classrooms within schools. We therefore cannot assume that each student's observations are completely independent of the groups that contain them. We adjust for this by estimating multivariate regressions on language arts and literacy, math, and science with cluster-adjusted standard errors. Similarly, effects on grade repetition and special education placement during elementary school were estimated using ordinary logistic regression analysis with cluster-adjusted standard errors as appropriate for a binary dependent variable.

All analyses include in addition to the central independent variable of interest (pre-K participation) the following control variables: student ethnicity, free or reduced-price lunch status, gender, age, employment status, home language, marital status, parental education, and school district. All models include district fixed effects (in essence children are compared only to other children within the same school district) where school district is the district in which the student entered kindergarten. In addition, we included dummy variables to denote the presence of missing values for each of the demographic variables obtained from the parent survey. Analyses were conducted separately for (a) the overall effects of any participation in Abbott Preschool and (b) the separate effects of one or two years of attendance in Abbott Preschool.

Fifth Grade Follow-up Findings

Table 3 presents estimates of the effects of the Abbott Preschool program on NJASK LAL, Math, and Science scores, retention, and special education placement after controlling for children's demographic characteristics. Effects on achievement test scores are reported for 4th and 5th grade. Grade retention is measured by whether a child is behind a grade level or more through grade 5 and special education is a measure of whether a child was in special education in

grade 4 or 5 (most often both). Children attending the Abbott Preschool Program gained an additional 5.0 points at 4th grade and 5.1 points at 5th grade on LAL, effect sizes of.16SD and .20SD, respectively. These gains correspond to an increase from the 50th percentile to the 58th for this measure of language abilities. In Math, Abbott pre-K attendees gained 9.0 points at 4th grade, an effect size of .24SD, and 7.6 points at 5th grade, an effect size of .19SD. In 4th grade Science, we found gains of 6.8 points, an effect size of .23SD. These effects correspond to an increase roughly from the 50th percentile to the 60th in the math and science achievement distribution. We also found that Abbott preschool attendance reduced the likelihood of in-grade retention by 40 percent and reduced the necessity of special education placement by 31 percent. These effects are statistically significant at the .05 level for the 5th grade tests and for the science test (available only for grade 4), and for the other measures at the .10 level (for a two-tailed test, which is conservative, and equivalent to a .05 significance level for a one-tailed test, which assumes negative effects are implausible).

Table 4 presents separate estimates of the effects of Abbott Pre-K for children who attended the program at age 3 and continued in the program for two years and for children who entered the program at age 4 and attended the program for one year. All of the estimated effects on test scores were numerically larger for two years of Abbott Pre-K (though differences between one- and two-year effects are not necessarily statistically significant differences), and, with one exception, they are significant at the .05 level. Effect sizes for two years of Abbott Pre-K range from .22SD to .37SD; for one year they range from .12SD to .18SD. This pattern of larger effect sizes for two years of Abbott Pre-K was not found for grade retention and special education. Differences in district policies and differences in the distribution of the one- and two-year samples by district may influence estimated effects on grade retention and special education

placement (for example, some districts have very minimal retention rates by policy). Table 5 shows the distribution of study participants by school district.

We conducted several additional analyses to assess the robustness of our findings to alternative analyses. Inclusion of a variable for comparison group participation in another preschool program had negligible impacts on the results. This could be because other programs were of low average quality, and errors in parental report (perhaps reporting family home day care or minimal attendance) could attenuate the effect of this variable, as well. We applied propensity score analysis to see if this might produce better estimates and found that it did not generate different results. Finally, we estimated all of the models without district fixed effects and again found that results differed little. We were particularly concerned that grade retention and special education placements were so rare in some districts that district fixed effects might produce more imprecise estimates. Estimated effects of pre-K on these outcomes were more often significant at the .05 level (two-tailed tests) without district fixed effects.

Discussion

The overall pattern of results in the APPLES 4th and 5th grade follow-up using New Jersey statewide assessment and school placement data provides strong evidence that the Abbott Preschool program has produced persistent, meaningful gains in achievement for children in the state's most disadvantaged communities. Achievement gains were particularly large for children who attended the program for two years. Substantive reductions in grade retention and special education placements were produced as well. These findings of persistent effects are all the more remarkable because the longitudinal study's approach had been found to underestimate the pre-K program's initial impact. Table 6 reports the estimates from the RDD kindergarten entry, APPLES kindergarten entry, APPLES 2nd grade, and the average APPLES effects across of 4th

and 5th grade for comparison over time. As expected, the effects at grades 4 and 5 for one year of pre-K are smaller than the initial effects estimated by the RDD study, as are the contemporaneous APPLES estimates at kindergarten entry. The effects of both one and two years of pre-K also tend to be somewhat smaller at grades 4 and 5 than found earlier for the APPLES sample. However, the effects on achievement remain substantial at 4th and 5th grade. As grade retention and special education are cumulative, reductions in those are somewhat larger in absolute terms at 5th grade than at 2nd grade.

When interpreting the APPLES estimates of persistent effects through 5th grade it is important to keep certain points in mind. First, it is likely that compensatory efforts by the public schools helped the children who did not attend Abbott Pre-K partially catch up to the others (Barnett, 2011). Such efforts are not free, and the higher percentages of no-pre-K children retained and placed in special education can be considered the most obvious, and costly, manifestations of that compensatory effort. Other efforts that schools commonly provide include reading tutors, in-class supports, and individualized lessons for selected students. Second, as children move through the grades the amount of progress made each year declines when measured as an effect size (Bloom, Hill, Black, & Lipsey, 2008). This means that apparently smaller effect sizes in 4th and 5th grades are not smaller than earlier gains when measured as a percentage of the average gain made in a year of school. To be more specific, by 5th grade the achievement advantage from attending two years of Abbott Pre-K still equals about three-quarters of a year of growth in math and two-thirds of a year of growth in language arts (allowing for underestimation, effects may be equivalent to more than a full grade level).

Whether the pre-K program's effects are interpreted as fractions of a year of growth or a percentage of the achievement gap (as discussed earlier), it is clear that pre-K is not a panacea.

Other efforts are needed to ensure that all children succeed throughout school and to more fully close the achievement gap for disadvantaged children. Some of New Jersey's Abbott districts have been noteworthy for their efforts in the years beyond pre-K and the results that they have produced (Kirp, 2013; MacInnes, 2009). High-quality pre-K is properly viewed as the leading edge of education reform, not as a complete solution on its own. Consideration should be given to the need to improve investments in child development in the years before age 3, as well.

Nevertheless, the value of the gains from Abbott Pre-K should not be minimized. The Abbott Preschool program's estimated long-term effects are considerably larger (by several multiples) than the estimated effects of less expensive state programs with lower standards (Fitzpatrick, 2008; Kuhne, 2008) and about 50 percent larger than for the most recent wave studied in the Tulsa, Oklahoma pre-K program (Hill, Gormley, & Adelstein, 2012). Although it is universal, the Tulsa program serves a population that is only slightly less disadvantaged than that in New Jersey's Abbott districts so the programs are quite comparable in this respect. The Tulsa program is relatively well-funded and has higher standards than many other state programs, but provides only one year of pre-K, has larger class sizes, and does not have the continuous improvement system that has been implemented for the Abbott Preschool program (Frede & Barnett, 2011). The continuous improvement system for Abbott Pre-K combines teacher self-assessment and collection of data on children's progress with coaching by master teachers to produce effective teaching; this may be one of the keys to the program's success (Pianta, Barnett, Burchinal, & Thornburg, 2009). As noted earlier, the APPLES study estimates the effects of Abbott Pre-K before this process had driven quality to current levels (as well as including only the largest districts). Therefore, children who entered Abbott Pre-K in more

recent years can be expected to have obtained even larger gains than those documented by this study.

The APPLES findings through 5th grade are best interpreted in the context of the larger body of knowledge regarding preschool education's effects reviewed earlier in this report. The Abbott Preschool program's effects on achievement and school success over time are consistent with the findings of research on preschool generally as summarized by meta-analysis. Abbott program effects also are similar in size to those produced by the Chicago Child Parent Centers. The Chicago program has been found to yield economic benefits that greatly exceed its costs (Reynolds, Temple, Robertson, & Mann, 2002). A substantial return on the state's investment might be expected for the Abbott program based on its comparable results so far. The Abbott program's effects are smaller than the effects of the most intensive (and expensive) model programs that have been studied (Campbell, Ramey, Pungello, Sparling, &Miller-Johnson, 2002; Schweinhart et al., 2005).

Broad gains in skills and knowledge, as well as reduced special education placement and grade retention, paint a consistent picture of increased school success from high-quality Abbott pre-K. The Abbott Preschool program's gains are practically meaningful and have already generated savings for taxpayers who pay less due to reductions in repeated years of schooling and special education that are no longer needed. A limitation of this study is that the state assessment data do not provide measures of social and emotional development, nor do we yet have measures beyond 5th grade. However, the larger literature strongly indicates that these results are predictive of such later outcomes as high school graduation and adult earnings, and that programs that address the needs of the whole child with a research-based curriculum produce broader improvements in such outcomes as delinquency and crime. This broad

approach is mandated by New Jersey's early learning standards and validated by direct observation of teaching that documents its implementation.

This study's findings add to the considerable body of evidence indicating that high-quality preschool education can significantly improve children's learning and development over the long term (Barnett, 2011; Burger, 2010; Camilli et al., 2010; Frede, 1998; Pianta et al., 2009). Moreover, this study specifically adds to the evidence that public pre-K programs can produce meaningful long-term educational gains on a large scale. In this context, our findings suggest that weak gains in the short and long term should be attributed to the failure to design (and fund) pre-K programs to replicate successful approaches in the first place rather than to problems of scaling up effective models. Although the Abbott Preschool program was implemented in communities that have relatively high percentages of low-income families for New Jersey, it is a universal program open to all within those communities and many of the children served are not from low-income families (30 percent did not qualify for free or reduced-price lunch). Therefore, the study also contributes important information to our knowledge of the impacts of universal pre-K.

Finally, in the context of the broader literature this study's findings have clear policy implications. New Jersey should stay the course with respect to ensuring that it maintains the quality of the Abbott Preschool program at high levels and that participation by children in the Abbott districts is maximized. In addition, as required by the New Jersey School Funding Reform Act of 2008, high-quality pre-K should be expanded to offer a comparable program to all low-income children. It is likely that New Jersey's citizens also would benefit from extending the opportunity for high-quality pre-K to all of the state's children, perhaps starting by offering one year at age 4. The state could facilitate this by extending the funding formula used for K-12

to include one year of pre-K for children beyond those in low-income families. Nationally, this study's findings support the wisdom of other states investing in high-quality pre-K and of the president's proposal in his 2013 State of the Union address to provide federal matching funds for state investments in high-quality pre-K. Finally, it adds to evidence that in addressing the needs of lower-income children, in particular, consideration should be given to providing more than one year of high-quality pre-K.

Table 1. Descriptive statistics for sample at kindergarten follow-up

	Abbott Pre-K (N=553)	Abbott Pre-K missing (N=213)	No Abbott Pre-K (N=201)	No Abbott Pre-K missing (N=104)
Demographic characteristics	Mean/%	Mean/%	Mean/%	Mean/%
Black	38.88%	44.60%	39.80%	41.35%
Hispanic	53.53%	46.95%	50.75%	50.96%
Other	2.53%	4.23%	3.48%	3.85%
Age (Months)	96.16 (3.71)	96.11 (3.84)	96.54 (3.89)	96.25 (4.09)
Gender (Male)	47.38%	53.52%	54.73%	54.81%
Parent Education (high school above)	20.61%	22.54%	15.92%	21.15%
Free/reduced lunch	69.44%	NA	77.11%	NA
Married family	32.01%	27.70%	30.35%	23.08%
Parental employment	43.40%	46.48%	36.32%	39.42%
Home language (English)	50.81%	56.81%	46.77%	52.88%
One year Pre K	59.49%	61.97%	NA	NA
Two year Pre K	40.51%	38.03%	NA	NA
K test scores	Mean (*SD*)	Mean (*SD*)	Mean (*SD*)	Mean (*SD*)
PPVT Fall	55.68 (17.43)	55.75 (17.25)	50.13 (21.18)	50.88 (18.36)
PPVT Spring	66.38 (17.12)	68.84 (16.34)	61.17 (19.18)	65.33 (15.89)
Woodcock-Johnson Fall	13.54 (4.24)	13.64 (4.10)	12.54 (4.44)	12.67 (4.25)
Woodcock-Johnson Spring	18.71 (5.25)	18.76 (4.14)	17.77 (4.85)	18.02 (4.59)
Print Awareness Fall	26.24 (8.88)	27.01 (8.43)	22.82 (9.82)	24.39 (9.25)
Print Awareness Spring	33.72 (3.86)	33.81 (3.94)	32.73 (5.11)	33.48 (4.24)

Table 2. Descriptive statistics for 5th grade follow-up sample

	Full Sample (N=754)	Abbott 1-year (N=329)	Abbott 2-year (N=224)	No Abbott pre-K (N=201)	Differences between groups
Demographic characteristics	Mean/%	Mean/%	Mean/%	Mean/%	
Black	39.12%	40.43%	36.61%	39.80%	
Hispanic	52.79%	51.37%	56.70%	50.75%	
Other	2.79%	2.47%	2.23%	3.48%	
Age (months)	108.26 (3.76)	96.20 (3.71)	96.11 (3.71)	108.54 (3.89)	
Gender (Male)	49.34%	48.63%	45.54%	54.73%	
Parental education (high school above)	19.37%	19.15%	22.77%	15.92%	
Free/reduced lunch	71.49%	70.52%	67.86%	77.11%	
Married family	31.57%	34.35%	28.57%	30.35%	
Parental employment	41.51%	41.95%	45.54%	36.32%	
Home language (English)	49.73%	48.94%	53.57%	46.77	
NJ ASK outcome measures	Mean (*SD*)	Mean (*SD*)	Mean (*SD*)	Mean (*SD*)	
Language Arts Literacy at 4th grade	187.67 (29.39)	187.39 (29.41)	193.30 (27.46)	181.95 (30.41)	**
Language Arts Literacy at 5th grade	187.37 (25.15)	188.56 (24.37)	190.28 (25.03)	182.29 (25.90)	**
Math at 4th grade	210.44 (38.45)	209.48 (37.53)	220.18 (38.46)	201.45 (37.71)	**
Math at 5th grade	213.80 (38.44)	213.75 (37.74)	220.74 (37.08)	206.46 (39.80)	**
Science	222.68 (28.72)	222.25 (28.95)	228.62 (27.15)	216.97 (28.92)	**
Retention	17.90%	14.89%	18.75%	21.89%	+
Special education	14.72%	12.16%	14.29%	19.40%	*

Significant Difference: ** $p < .01$, * $p < .05$, + $p < .10$

Table 3. Regression analysis estimating effects of Abbott Pre-K program on NJ ASK

	LAL 4th	LAL 5th	Math 4th	Math 5th	Science	Retention	Special ed.
Abbott pre-K	5.013+	5.052*	8.951*	7.627*	6.788**	-.495+	-.365+
	(2.594)	(2.464)	(3.530)	(3.805)	(2.506)	(.269)	(.228)
Free/Reduced lunch	-5.748+	-6.314*	-9.242*	-7.109+	-7.828**	.724*	.238
	(2.955)	(2.494)	(3.994)	(4.123)	(2.999)	(.324)	(.304)
Age	.773**	.678*	1.296**	.804+	1.092**	-.131**	-.003
	(.276)	(.297)	(.343)	(.456)	(.283)	(.033)	(.033)
Married two parents family	1.602	-2.083	1.748	1.938	2.859	-.118	-.353
	(3.027)	(2.401)	(3.811)	(3.818)	(2.877)	(.285)	(.369)
Employment	.131	-2.759	3.502	-2.423	-1.374	.214	-.075
	(2.338)	(2.344)	(4.005)	(3.934)	(3.058)	(.226)	(.270)
Parent educational level	1.374	2.531+	1.003	3.349	1.757	-.112	-.024
	(1.794)	(1.341)	(2.371)	(2.252)	(1.828)	(.136)	(.180)
Home Language (English=1)	2.077	.395	-3.936	-1.046	-.394	.295	-.177
	(3.178)	(2.824)	(4.166)	(3.795)	(3.180)	(.266)	(.328)
Gender (female=1)	8.302**	6.389**	.106	3.674	1.738	-.768**	-1.309**
	(2.200)	(1.863)	(2.638)	(2.997)	(2.133)	(.209)	(.252)
Black	-13.270*	-13.602*	-21.818**	-28.260**	-11.880+	.058	.360
	(5.987)	(6.465)	(8.018)	(7.760)	(6.351)	(.493)	(.485)
Hispanic	-3.353	-3.230	-10.540	-9.329	-2.082	.157	-.185
	(5.698)	(6.700)	(8.365)	(7.805)	(6.072)	(.451)	(.521)
Other	10.689	4.516	13.816	9.358	8.072	.216	-.800
	(8.028)	(9.364)	(12.595)	(12.203)	(9.005)	(.720)	(1.228)
N	658	661	658	662	637	743	644
R-square	.17	.15	.17	.16	.15	.13	.10
Effect size	.16+	.20*	.24*	.19*	.23**	40%+	31%+

Significant Effect: ** $p < .01$, * $p < .05$, + $p < .10$

Table 4. Regression analysis estimating effects of Abbott Pre-K program by years of participation

	LAL 4th	LAL 5th	Math 4th	Math 5th	Science	Retention	Special ed.
One year of	3.569	4.785*	6.572+	5.743	4.884+	-.621*	-.438+
Abbott pre-K	(2.637)	(2.388)	(3.504)	(3.677)	(2.477)	(.276)	(.251)
Two years of	8.003*	5.609+	13.892**	11.531*	10.711**	-.275	-.224
Abbott pre-K	(3.099)	(3.129)	(4.484)	(4.763)	(3.390)	(.321)	(.280)
Free/Reduced lunch	-5.565+	-6.293*	-8.935*	-6.962+	-7.630*	.736*	.246
	(2.945)	(2.495)	(3.978)	(4.088)	(2.986)	(.324)	(.307)
Age	.773**	.676*	1.293**	.794+	1.087**	-.131**	-.003
	(.276)	(.298)	(.343)	(.458)	(.287)	(.032)	(.032)
Married two parents family	1.947	-2.027	2.323	2.330	3.228	-.084	-.331
	(3.044)	(2.417)	(3.747)	(3.772)	(2.873)	(.292)	(.369)
Employment	.128	-2.761	3.498	-2.432	-1.404	.212	-.070
	(2.335)	(2.351)	(3.951)	(3.905)	(3.068)	(.227)	(.272)
Parent educational level	1.356	2.525+	.971	3.305	1.681	-.102	-.019
	(1.798)	(1.344)	(2.354)	(2.239)	(1.844)	(.135)	(.179)
Home language	1.625	.307	-4.683	-1.658	-1.016	.259	-.195
(English=1)	(3.146)	(2.877)	(4.155)	(3.806)	(3.132)	(.270)	(.330)
Gender	8.300**	6.388**	.121	3.667	1.733	-.770**	-1.311**
(female=1)	(2.192)	(1.861)	(2.632)	(2.992)	(2.140)	(.207)	(.252)
Black	-13.706*	-13.704*	-22.510**	-28.974**	-12.607*	.006	.339
	(5.964)	(6.453)	(7.971)	(7.792)	(6.371)	(.495)	(.488)
Hispanic	-3.900	-3.344	-11.448	-10.125	-2.855	.104	-.210
	(5.631)	(6.686)	(8.290)	(7.855)	(6.009)	(.448)	(.523)
Other	9.832	4.343	12.384	8.148	6.887	.157	-.822
	(8.113)	(9.375)	(12.864)	(12.395)	(9.105)	(.716)	(1.236)
N	658	661	658	662	637	743	644
R-square	.17	.15	.18	.16	.16	.14	.10
Effect size (Abbott 1)	.12	.18*	.17+	.14	.17+	47%*	36%+
Effect size (Abbott 2)	.26*	.22+	.37**	.29*	.37**	35%	21%

Significant Effect: ** $p < .01$, * $p < .05$, + $p < .10$

Table 5. Follow-up sample size by school district

District	Follow up Total (N=754)	Abbott Pre-K (N=553)	No Abbott Pre-K (N=201)
Camden	60 (7.96%)	28 (5.25%)	32 (15.92%)
East Orange	46 (6.10%)	42 (7.88%)	4 (1.99%)
Elizabeth	62 (8.22%)	48 (9.01%)	14 (6.97%)
Irvington	24 (3.18%)	18 (3.38%)	6 (2.99%)
Jersey City	51 (6.76%)	40 (7.50%)	11 (5.47%)
New Brunswick	21 (2.79%)	12 (2.25%)	9 (4.48%)
Newark	103 (13.66%)	63 (11.82%)	40 (19.90%)
Passaic	67 (8.89%)	56 (10.51%)	11 (5.47%)
Paterson	106 (14.06%)	78 (14.63%)	28 (13.93%)
Perth Amboy	16 (2.12%)	11 (2.06%)	5 (2.49%)
Plainfield	10 (1.33%)	9 (1.69%)	1 (.50%)
Trenton	64 (8.49%)	46 (8.63%)	18 (8.96%)
Union City	47 (6.23%)	38 (7.13%)	9 (4.48%)
Vineland	43 (5.70%)	35 (6.57%)	8 (3.98%)
West New York	34 (4.51%)	29 (5.44%)	5 (2.49%)

Table 6. Abbott Pre-K effect sizes over time

	APPLES			
	RDD	K Entry	2nd	4th/5th
PPVT				
Year 1	.28	.21	.22	NA
Year 2	NA	.42	.40	NA
Mathematics				
Year 1	.36	.20	.24	.17 [c]
Year 2	NA	.34	.44	.37 [c]
Literacy/Language Arts				
Year 1	.56 [a]	.29 [a]	.16 [b]	.15 [d]
Year 2	NA	.31 [a]	.20 [b]	.24 [d]

Note. [a] Kindergarten literacy was measured using the Pre-CTOPP.
[b] Second grade literacy was measured by WJ subtest Passage Comprehension.
[c] NJ-Ask 4th and 5th grade average of effects mathematics.
[d] NJ-Ask 4th and 5th grade average of effects on language arts and literacy.

References

Barnett, W. S. (2011). Effectiveness of early educational intervention. *Science, 333*, 975-978.

Barnett, W. S., Epstein, D. J., Carolan, M. E., Fitzgerald, J., Ackerman, D. J., & Friedman, A. H. (2010). *The state of preschool 2010: State preschool yearbook.* New Brunswick, NJ: National Institute for Early Education Research, Rutgers University.

Barnett, W. S., Howes, C., & Jung, K. (2008). *California's state preschool program: Quality and effects on children's cognitive abilities at kindergarten entry.* New Brunswick, NJ: National Institute for Early Education Research, Rutgers University.

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, *1*(4), 289-328.

Burger, K. (2010). How does early childhood care and education affect cognitive development? An international review of the effects of early interventions for children from different social backgrounds. *Early Childhood Research Quarterly*, *25*, 140-165.

Camilli, G., Vargas, S., Ryan, S., & Barnett, W.S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record, 112*(3), 579-620.

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.

Consortium for Longitudinal Studies (Eds.) (1983). *As the twig is bent…lasting effects of preschool programs.* Hillsdale, NJ: Erlbaum.

Cook, T.D. (2008). Waiting for life to arrive: A history of the regression-discontinuity design in psychology, statistics, and economics. *Journal of Econometrics, 142*(2), 636-654.

Deutsch, M., Deutsch, C. P., Jordan, T. J., & Grallow, R. (1983). The IDS program: An experiment in early and sustained enrichment. In Consortium for Longitudinal Studies (Eds.), *As the twig is bent… Lasting effects of preschool programs* (pp. 377-410). Hillsdale, NJ: Lawrence Erlbaum.

Deutsch, M., Taleporos, E., & Victor, J. (1974). A brief synopsis of an initial enrichment program in early childhood. In S. Ryan (Ed.), *A report on longitudinal evaluations of preschool programs Volume 1: Longitudinal evaluations* (pp. 49-60). Washington, DC: Office of Child Development, US Department of Health, Education, and Welfare.

Fitzpatrick, M. D. (2008). Starting school at four: The effect of universal pre-kindergarten on children's academic achievement. *The BE Journal of Economic Analysis & Policy, 8*(1), Article 46.

Frede, E. C. (1998). Preschool program quality in programs for children in poverty. In W. S. Barnett & S. S. Boocock (Eds.), *Early care and education for children in poverty* (pp. 77-98). Albany, NY: SUNY Press.

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.

Frede, E. C., & Barnett, W. S. (2011). New Jersey's Abbott pre-k program: A model for the nation. In E. Zigler, W. Gilliam, & W. S. Barnett (Eds.), *The pre-k debates: current controversies and issues* (pp. 191-196). Baltimore, MD: Brookes 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., Jung, K., Barnett, W. S., & Figueras, A. (2009). *The APPLES blossom: Abbott Preschool Program Longitudinal Effects Study (APPLES), preliminary results through 2nd grade*. New Brunswick, NJ: National Institute for Early Education Research, Rutgers University.

Gormley, W. T., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-k on cognitive development. *Developmental Psychology, 41*(6), 872-884.

Gormley, W. T., Phillips, D., & Gayer, T. (2008). Preschool programs can boost school readiness. *Science, 320*, 1723-24.

Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early childhood environmental rating scale-revised edition*. New York: Teachers College Press.

Hill, C., Gormley, W., & Adelstein, S. (2012). Do the short-term effects of a strong preschool program persist? Working Paper # 18 Washington, DC: CROCUS, Georgetown University.

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, Rutgers University.

Hustedt, J. T., Barnett, W. S., Jung, K., & Thomas, J. (2007). *The effects of the Arkansas Better Chance Program on young children's school readiness*. New Brunswick, NJ: National Institute for Early Education Research, Rutgers University.

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, 74-80.

Jordan, T. J., Grallo, R., Deutsch, M., & Deutsch, C. P. (1985). Long-term effects of early enrichment: A 20-year perspective on persistence and change. *American Journal of Community Psychology, 13*(4), 393-415.

Kirp, D. L. (2013). *Improbable Scholars: The Rebirth of a Great American School System and a Strategy for America's Schools*. OUP USA.

Kuhne, K.A. (2008). Pre-k effects in Texas (Dissertation). Dallas: University of Texas at Dallas.

Lamy, C., Frede, E., & ELIC. (2005). *Giant steps for the littlest children: Progress in the sixth year of the Abbott Preschool Program*. Trenton, NJ: New Jersey Department of Education. Available online at www.nj.gov/njded/ece

Larsen, J. M., & Robinson, C. C. (1989). Later effects of preschool on low-risk children. *Early Childhood Research Quarterly*, *4*, 133–144.

Lipsey, M. W., Farran, D. C., Hofer, K. G., Bilbrey, C., & Dong, N. (2011) *The effects of the Tennessee voluntary pre-kindergarten program: Initial result.* Nashville, TN: Peabody Research Institute, Vanderbilt University.

MacInnes, G. (2009). *In Plain Sight: Simple, Difficult Lessons from New Jersey's Expensive Efforts to Close the Achievement Gap*. Century Foundation Press.

Malofeeva, E., Daniel-Echol, M., & Xiang, Z. (2007). *Findings from the Michigan School Readiness Program 6 to 8 follow up study.* Yspsilanti, MI: High Scope Educational Research Foundation.

NJ DOE. (2002a). *New Jersey kindergarten core curriculum content standards.* Trenton, NJ: Author.

NJ DOE. (2002b). *New Jersey preschool teaching and learning expectations: Standards of quality*. Trenton, NJ: Author.

NJ DOE. (2005). *Abbott Preschool Program implementation guidelines* (Rev. ed.). Trenton, NJ: Author.

Peisner-Feinberg, E. S., & Schaaf, J. M. (2010). *Long-term effects of the North Carolina More at Four Pre-Kindergarten Program: Children's reading and math skills at third grade.* Chapel Hill, NC: Frank Porter Graham Child Development Institute, University of North Carolina.

Pianta, R. C., Barnett, W. S., Burchinal, M., & Thornburg, K. (2009). The effects of preschool education: What we know, how public policy is or is not aligned with the evidence base, and what we need to know. *Psychological Science in the Public Interest, 10*(2), 49-88.

Reynolds, A. J. (2000). *Success in early intervention: The Chicago Child-Parent Centers*. Lincoln: University of Nebraska Press.

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. *Educational Evaluation and Policy Analysis, 24*, 267-303.

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.

Schweinhart, L.J., Xiang, Z., Daniel-Echols, M., Browning, K. & Wakabayashi, T. (2012). *Michigan Great Start Readiness Program evaluation 2012: High school graduation and grade retention findings*. Ypsilanti, MI: HighScope Educational Research Foundation

Temple, J., & Reynolds, A. (2007). Benefits and costs of investments in preschool education: Evidence from the Child-Parent Centers and related programs. *Economics of Education Review, 26,* 126-144.

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

U.S. Department of Health and Human Services, Administration for Children and Families (2010). *Head Start Impact Study. Final report*. Washington, DC.

Whitehurst, G.J. (2013). Can we be hard-headed about preschool? A look at universal and targeted pre-K. Downloaded March 11, 2013 from http://www.brookings.edu/blogs/brown-center-chalkboard/posts/2013/01/16-preschool-whitehurst

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