CEPP

Discussion Paper Series

**Investing in Preschool Programs**

Greg J. Duncan
University of California, Irvine

Katherine Magnuson
University of Wisconsin - Madison

Center for Economics & Public Policy | University of California, Irvine



www.economicsandpublicpolicy.uci.edu | 949.824.8496 | dneumark@uci.edu
3151 Social Science Plaza | Irvine, CA 92697-5100

## Executive Summary

## Investing in Preschool Programs

At the beginning of kindergarten, the math and reading achievement gaps between children in the bottom and top income quintiles amount to the equivalent of over 100 SAT points -- more than a full standard deviation. Early childhood education programs provide child care services and may facilitate the labor market careers of parents, but their greatest potential value is as a human capital investment in young children, particularly children from economically disadvantaged families. Can early childhood education programs can be designed to provide the kinds of enrichment that low-income children most need to do well in school and succeed in the labor market?

In this study, Greg Duncan and Katherine Magnuson summarize the available evidence on the extent to which expenditures on early childhood education programs constitute worthy social investments in the human capital of children. A brief summary of their analysis is presented here.

### Analysis

*Meta-Analysis: Comparing Results across Studies*

A comparison of the cognitive and achievement treatment effects from 84 early childhood education programs shows the following:

- The simple average effect size across all 84 programs is .35 standard deviations, an amount equal to nearly half of the racial achievement gap for kindergarten-aged kids.
- The studies with the fewest subjects (and consequently least amount of precision) had the largest effect sizes - when the precision of each study is taken into account, the average effect size drops to .21 standard deviations.
- Programs beginning before 1980 produced significantly larger effect sizes (.33 standard deviations) than those that began later (.16 standard deviations). While declining effect sizes over time are disappointing, they are likely driven by the fact that, over time, the quality of the home environments provided by low-income families has improved and many more low-income children are enrolled in child care centers than before.
- Longer-duration programs do not appear to have larger effects.
- Evidence on starting age is mixed, but taken as whole, effect sizes were neither larger nor smaller for children who started programs at younger ages.

*Model Program Impacts: Perry Preschool and Abecedarian*

The Perry Preschool and Abecedarian programs are prominent studies from the 1960s and 1970s. While achievement gains for these programs, as measured by test scores, tend to fade over time, a key reason for the importance of these studies is that long-term follow-ups show strikingly positive impacts in adulthood. Nevertheless, it is difficult to extract policy lessons from these initiatives for early childhood education programs that might be offered by government today. Both programs were designed and evaluated by researchers and each served only several dozen children – conditions that scaled-up programs cannot match. In addition, the low levels of maternal education for both Perry and Abecedarian treatment groups may have led to larger program impacts than we would expect to achieve today given the relative rise in years of completed schooling by poor mothers.

*Head Start Impacts*

Large-scale policy lessons might be gleaned more reliably from studies of Head Start, since that program now provides services to almost a million three- and four-year-olds. Both quasi-experimental and a random-assignment evaluations of Head Start found significant short-term gains in participants' achievement test scores but, as with Perry and Abecedarian, these gains appeared to fade over time. However, as with Perry and Abecedarian programs, quickly declining test score impacts for recent cohorts of Head Start children appear to be at odds with the long-term impacts on important young adult outcomes found in analyses of older Head Start cohorts. Some of the older-cohort studies use strong quasi-experimental methods and find quite striking long-run program impacts.

*Pre-Kindergarten Programs*

Some rigorous evaluations of pre-kindergarten programs were completed too recently to have been included in the meta-analysis described earlier. Several of these studies find short-run effects on achievement test scores that are somewhat larger than those estimated in the national Head Start Impact Study. However, longer-run impacts are only evaluated up to the third grade. While there is evidence of persisting math impacts (.18 standard deviations), the lack of longer-run evaluations beyond the third grade suggests that drawing strong policy conclusions about their effectiveness is unwarranted. Other programs have likewise demonstrated early promising results that faded over the first few years of school.

*The Puzzle: Academic Fade-Out of Long-Term Benefits*

Most early childhood education studies that have tracked children beyond the end of the program treatment find that effects on test scores fade over time. An analysis of cognitive and achievement outcomes in the meta-analytic database shows an estimated decrease in program impact effect sizes of about .03 standard deviations per year, which implies that positive effects persist for roughly 10 years. This finding raises a puzzle: How do we reconcile the fade-out of preschool program impacts on test scores during elementary school with the evidence showing that such programs nonetheless have beneficial impacts on a broad set of later-life outcomes like high school graduation rates, teen parenthood, and criminality? One possible answer is that preschool programs may affect something other than basic achievement and cognitive test scores, and perhaps these other program impacts, unlike achievement and cognitive impacts, persist over time. Several theories of human development that allow for preschool programs to generate more broad impact on children's behavior and social competence have been presented. Despite arguments in their favor, testing these theories is difficult since most preschool studies do not measure many of these kinds of outcomes at program completion. Those studies that have included measures of problem behavior have produced mixed results. Reconciling disparate patterns of impacts in the short and longer term is a key challenge for anyone hoping to extract policy lessons about the effectiveness of early childhood education programs.

*Within-Program Heterogeneity*

Although policymakers appropriately care most about the average impacts of early childhood education programs, a number of lessons can be learned from looking at how the effects vary across certain sub-groups. A finding of heterogeneous impacts might make it possible to identify groups that could particularly benefit from the preschool setting. However, a systematic accounting of heterogeneity in the effects of preschool programs is a complicated undertaking. Even when studies determine that a particular program has been a success overall, the positive outcomes differ across programs and populations. Efforts to accurately identify differential effects across subgroups are often hampered by small sample sizes. Nonetheless, given the potential benefits, greater attention should be give to understanding both who benefits the most from particular programs and why.

*The Search for Active Program Ingredients*

Research on early childhood education has focused greater attention on evaluating particular programs than on identifying the particular ingredients in these programs that produce significant improvements in children's learning and behavior. Some scholars have focused on structural aspects of early childhood education environments, such as class size and teacher education, yet these features of programs are likely to affect children only indirectly, by influencing their experiences within classrooms. Potentially more important for children's actual experiences in early childhood education programs, is what developmental psychologists have referred to as "process quality" – the quality of classroom interactions, including the amount of instructional and emotional support children receive. As attention has shifted to improving classroom interactions, two aspects of program design emerge as policy levers that may, together, improve program effectiveness: curriculum and related professional development.  It appears that an effective strategy is to combine a proven curriculum that offers well-designed lesson plans and activities based on an understanding of children's trajectories of learning within specific content areas, with strong professional development to target improvement in specific instructional practices.

**Policy Recommendations**

Evaluations of most early childhood education programs show that they improve children's school readiness. Some longer-run evidence shows that these impacts can extent into adulthood, although it is unclear whether this is likely to be the case with large-scale programs offered today. Most promising today are pre-kindergarten classrooms with proven curricula and well-trained staff.

More specific policy recommendations require a change in how research is conducted in this area. Rather than looking merely at average short-run outcomes of early childhood education programs based on a limited number of achievement tests, researchers should focus on the heterogeneity of outcomes across groups, conduct long-term follow-up, and examine a wide range of outcome variables that would illuminate the program ingredients and developmental processes that make some of these programs so successful.

**Investing in Preschool Programs**

At the beginning of kindergarten, the math and reading achievement gaps between children in the bottom and top income quintiles amount to more than a full standard deviation. Early childhood education programs provide child care services and may facilitate the labor market careers of parents, but their greatest potential value is as a human capital investment in young children, particularly children from economically disadvantaged families (Heckman, 2006). After all, both human and animal studies highlight the critical importance of experiences in the earliest years of life for establishing the brain architecture that will shape future cognitive, social, and emotional development, as well as physical and mental health (Sapolsky 2004; Knudsen et al. 2006).  And research on the malleability (plasticity) of cognitive abilities finds these skills to be highly responsive to environmental enrichment during the early childhood period (Nelson and Sheridan, 2011).  Perhaps early childhood education programs can be designed to provide the kinds of enrichment that low-income children most need to do well in school and succeed in the labor market.

We summarize the available evidence on the extent to which expenditures on early childhood education programs constitute worthy social investments in the human capital of children.  We begin with a short overview of existing early childhood education programs, and then summarize results from a substantial body of methodologically sound evaluations of the impacts of early childhood education. We find that the evidence supports few unqualified conclusions. Many early childhood education programs appear to boost cognitive ability and early school achievement in the short run. However, most of them show smaller impacts than those generated by the best known programs, and their cognitive impacts largely disappear within a few years. Despite this fade-out, long-run follow-ups from a handful of well-known programs show lasting positive effects on such outcomes as greater educational attainment, higher earnings, and lower rates of crime. Since findings regarding short and longer-run impacts on “noncognitive” outcomes are mixed, it is uncertain what skills, behaviors, or developmental processes are particularly important in producing these longer run impacts.

Our review also describes different models of human development used by social scientists, examines heterogeneous results across groups and tries to identify the ingredients of early childhood education programs that are most likely to improve the performance of these programs.  We use the terms “early childhood education” and “preschool” interchangeably to denote the subset of programs that provide group-based care in a center setting and offer some kind of developmental and educational focus. This definition is intentionally broad, as historical distinctions between early education and other kinds of center-based child care programs have blurred. Many early education programs now claim the dual goals of supporting working families and providing enriched learning environments to children, while many child care centers also foster early learning and development (Adams and Rohacek 2002).

**Existing Preschool Programs**

Most children enrolled in early childhood education attend private programs, some non-profit and others for-profit. In 2011, the average cost of full-time center-based care for a four-year old ranged from $3,900 in Mississippi to just over $14,000 in the District of Columbia (National Association of Child Care Resource and Referral Agencies 2011). Given the high cost of care, it is unsurprising that enrollment rates of children residing in families with incomes in the bottom half of the income distribution are persistently10-20 percentage points lower than for children in the highest quarter.   Figure 1, based on the data from the October Supplement to the

Current Population Survey, shows this enrollment gap by income level. The figure also shows a steady rise in enrollment in early childhood education programs among three- and four year-olds over the past 50 years. This increase is broad-based: across income groups and for the children of both employed and non-employed mothers.

States and the federal government have sought to increase the participation of low-income children in early childhood education programs in a number of ways: through Head Start, pre-kindergarten programs, and means-tested child-care assistance programs that can be used to pay for center-based care.[1] Overall, both federal and state investments in these programs increased substantially in real terms through the early 2000s, but in more recent years funding has not grown substantially (Barnett et al. 2011; Magnuson & Shager, 2010; Schulman & Blank, 2012).

*Head Start*, the federal government's largest compensatory preschool program, is designed to enhance children's social and cognitive development by providing a comprehensive set of educational, health, nutritional, and other social services. In 2005, virtually all Head Start programs were center-based and half offered full-day (six hours or more) services, five days a week (Hamm, 2006). Most children enrolled in Head Start in 2009 were three (36 percent) or four years old (51 percent). In 2010, the federal Head Start appropriation of about $7.2 billion was distributed to 1,591 local private and public non-profit grantees serving 904,153 children. Some states supplement federal funds to increase access to Head Start programs; for details, see the Head Start website at http://eclkc.ohs.acf.hhs.gov/hslc/mr/factsheets/fHeadStartProgr.htm. Local grantees are required to provide at least 20 percent matching funds which brings program costs to around $9,000 per child per year (Ludwig and Phillips, 2007).

*Pre-kindergarten programs* are funded primarily by states or local school districts. In 2011, 39 states and the District of Columbia spent about $5.5 billion on pre-kindergarten initiatives that collectively served approximately 28 percent of the nation's four-year-olds and 4 percent of three-year-olds (for details, see Barnett, Carolan, Fitzgerald, and Squires 2011). Most pre-kindergarten programs target low-income children (31 state programs have income eligibility requirements), and most offer health, vision, and hearing screenings as well as at least one other form of support service. One-half of state pre-kindergarten programs require teachers to have training in early child development and nearly one-third require BA degrees. Typically, states use a mixed service delivery system that provides programming in local elementary schools as well as community-based settings.

With expenditures in 2010 amounting to approximately $9.5 billion, federal and state funded *means-tested child care subsidies* can be used for various types of child care, including center-based care, family day care, and other forms of informal care, and they cover a wide age range of children (birth through age 12). Their primary goal has continued to be supporting working families rather than educating young children, although increased spending on subsidies has been linked to higher rates of preschool attendance among young children (Magnuson, Meyers, and Waldfogel 2007). Because parents' preferences and needs for child care may not always align well with what is provided by preschool programs, and because child care subsidy spells are often quite short (Ha, Magnuson and Ybara 2012), these subsidies are best viewed as an indirect way to promote early childhood education for three- and four-year-olds.

**Empirical Studies of the Effectiveness of Early Childhood Education**

Empirical studies of the effects of investments in early childhood education on children's human capital encompass a range of methodologies and a wide variety of programs. We focus on evaluations of preschool programs conducted over the course of the last half-century that are based on strong experimental or quasi-experimental methods and provide impact estimates for cognitive or achievement-related outcomes (see online appendix for details).[2] Despite the hundreds of evaluation studies of early childhood education programs that have been published over the past 50 years, a handful of programs have figured especially prominently in policy discussions: in particular Perry Preschool, the Abecedarian program, Head Start, and more recently some state and local pre-kindergarten programs.

*Meta-Analysis*

Figure 2 shows the distribution of 84 program-average treatment effect sizes for cognitive and achievement outcomes, measured at the end of each program's treatment period, by the calendar year in which the program began. Reflecting their approximate contributions to weighted results, "bubble" sizes are proportional to the inverse of the squared standard error of the estimated program impact. The figure differentiates between evaluations of Head Start and other ECE programs and also includes a weighted regression line of effect size by calendar year.

Taken as a whole, the simple average effect size for early childhood educations on cognitive and achievement scores was .35 standard deviations at the end of the treatment periods, an amount equal to nearly half of race differences in the kindergarten achievement gap (Duncan and Magnuson 2011). However, as can be seen from Figure 2, average effect sizes vary substantially and studies with the largest effect sizes tended to have the fewest subjects. When weighted by the inverse of the squared standard errors of the estimates, the average drops to .21 standard deviations.

All of the 84 programs that generated the effect size data shown in Figure 2 met minimum standards for quality of research methods. However, some of the programs lasted for only a couple of summer months, while others ran for as long as five years. Some of the evaluations used random assignment while others relied on less rigorous quasi-experimental methods. Almost all focused on children from low-income families, but they varied in the racial and ethnic composition of treatment groups.

One might assume that these differences would account for much of the effect-size variability observed in Figure 2. However, that is not always the case. Weighted average effect sizes were insignificantly different between evaluations that did (.25 standard deviations) and did not (.19 standard deviations) use random assignment; that were (.31 standard deviations) and were not (.18 standard deviations) published in peer-review journals. The effect sizes of programs designed by researchers (.39 standard deviations) were significantly larger than programs not designed by researchers (.18 standard deviations).

Programs beginning before 1980 produced significantly larger effect sizes (.33 standard deviations) than those that began later (.16 standard deviations). Declining effect sizes over time are disappointing, as we might hope that lessons from prior evaluations and advances in the science of child development would have led to an increase in program impacts over time. However, the likely reason for the decline is that counterfactual conditions for children in the control groups in these studies have improved substantially. We have already seen in Figure 1 how much more likely low-income children are to be attending some form of center-based care now, relative to 40 years ago. This matters because although center-based care programs have varying degrees of educational focus, most research suggests that center-based care is associated with better cognitive and achievement outcomes for preschool age children (NICHD and Duncan 2003).

Even more impressive are gains in the likely quality of the home environment provided by low-income mothers, as indexed by their completed schooling. In 1970, some 71 percent of preschool age children in the bottom 20 percent of the income distribution had mothers who lacked a high school degree, while only 5 percent of the mothers had attended at least some post-secondary schooling (based on authors' calculation of the October Current Population Survey data). By 2000, the corresponding percentage of children with mothers who did not have a high school degree had dropped by nearly half (to 37 percent), while the percentage with mothers who had completed some post-secondary schooling increased five-fold (to just over 25 percent). Today, therefore, children from low-income households are likely to be benefiting from much higher-quality home environments than their counterparts four decades ago. Both higher-quality home environments and increases in other forms of center-based child care raise the bar for impact estimates coming from early childhood education programs.

Two particularly salient features of early childhood education programs are duration and starting age. Abundant literature suggests that the number of years spent in K-12 or post-secondary education is linked to labor market success (Card 1999). Thus, it seems plausible to expect that longer exposure to early childhood education environments before school entry should boost later academic achievement as well. But while simple associations indicate that longer participation in a preschool program generates larger treatment effects, models with a full set of controls for program and evaluation quality yield only small and statistically insignificant associations (+.04 sd per additional year) between program duration and magnitudes of impacts

(Leak et al. 2012). The absence of larger effects for longer-duration programs may be due to an inability of such programs to use curricula and activities that are designed to capitalize fully on the skills gained in the early years of program participation.

As for starting age, neuroscience evidence on the plasticity of cognitive and language abilities suggests that these skills are highly amenable to environmental enrichment during the early childhood period. Starting in infancy, responsive caregiving and language-rich interactions are associated with better developmental outcomes, and more specifically stronger early language development (Tamis-LeMonda, Bornstein, and Baumwell 2001). Based on such findings, we might expect to find an "earlier is better" pattern of effects for if early childhood education programs that provide such high-quality interactions for children. Evidence from the best-known early-life preschool programs is mixed; programs such as Early Head Start produce very small impacts on cognitive development (Love et al. 2003), whereas others, like the Abecedarian program, show much larger impacts (Ramey and Campbell 1984). Analysis of the meta-analytic database shows that, taken as a whole, effect sizes were neither larger nor smaller for children who started programs at younger ages (Leak et al. 2012). This suggests that other modes of early childhood investments for example, home visitation for high-risk, first-time mothers (Olds, Sadler and Kitzman 2007) or developmental screenings and interventions for children living in families with documented domestic violence—may be more effective ways of building children's capacities during the very early years of life.

*Model Program Impacts: Perry Preschool and Abecedarian*

As shown in Figure 2, average end-of-treatment effect sizes for the Perry Preschool and Abecedarian programs are several times larger than the weighted mean effect size for all studies in the meta-analytic database that met our inclusions criteria. A key reason for the prominence of some of these two studies is that long-term follow-ups show strikingly positive impacts in adulthood and impressive benefit-cost ratios.

Perry provided one or two years of part-day educational services and weekly home visits to 58 low-income, low-IQ African American children aged three and four in Ypsilanti, Michigan, during the 1960s. The curriculum was geared to the children’s age and capabilities, emphasizing child-initiated learning activities. Staff encouraged children to engage in play activities that would promote their problem-solving skills as well as their intellectual, social, and physical development. Program staff made weekly one- to two-hour afternoon visits to each family. The center’s child-to-teacher ratio was low; each of four teachers served only 20-25 children every year. Per-pupil costs amounted to about $20,000 per child (in 2011 dollars). While Perry’s large impacts on IQ at the point of school entry had all but disappeared by third grade (Schweinhart et al. 2005), the program produced lasting improvements through age 40 on employment rates and substantially reduced the likelihood that participants had been arrested. Heckman and colleagues (2010) estimate that the program generated about $152,000 in benefits over the life course, boosting individuals’ earnings, reducing use of welfare programs and, most importantly for the benefit calculation, reducing criminal activity. These financial benefits produced a social rate of return between 7 and 10 percent.

The Abecedarian program, which served 57 low-income, mostly African American families from Chapel Hill, North Carolina, provided even more intensive services than Perry Preschool. Beginning in 1972, children assigned to the Abecedarian “treatment” received year-round, full-time center-based care for five years, starting in the child’s first year of life. The Abecedarian preschool program included transportation, individualized educational activities that changed as the children grew older, and low child-teacher ratios of 3:1 for the youngest children and up to 6:1 for older children. Abecedarian teachers followed a curriculum that focused on language development and explained to teachers the importance of each task as well as how to teach it. High-quality health care, additional social services, and nutritional supplements were also provided to participating families (Ramey and Campbell 1979; Campbell et al. 2002).

At two years of age, the control-group children in the Abecedarian program had IQ scores that averaged about one standard deviation below the mean, as would be expected for children from very economically disadvantaged backgrounds (Ramey, Campbell, Burchinal, Skinner, Gardner, and Ramey 2000). By the time the children reached age five, however, their IQ scores were close to the national average, and 10 points higher than scores of comparable children who did not participate in the program. Similarly large effects were observed for achievement on verbal and quantitative tests (Ramey and Campbell 1984). Nearly 15 years later, the program's effect on IQ scores at age 21 (.38 standard deviations) was still substantial but smaller than at age five. Children in the Abecedarian program entered college at 2.5 times the rate of children in the control group, and the intervention also reduced rates of teen parenthood and marijuana use by nearly half, although it did not lead to statistically significant reductions in criminal activity. Expressed in 2011 dollars, the costs associated with Abecedarian's five-year duration totaled about $80,000 per child, and the program is estimated to have produced $160,000 in net present benefits for its participants and their parents (Barnett and Masse 2007; Currie 2001).

It is difficult to extract policy lessons from these initiatives for early childhood education programs that states or the federal governments might offer today. Both programs were designed and evaluated by researchers and each served only several dozen children – conditions that scaled-up programs cannot match. Moreover, as we have pointed out above, counterfactual conditions three decades ago were likely of a comparatively low quality. The average number of years of maternal education completed was about 10 years for both the Perry and Abecedarian preschool treatment groups, reflecting the low levels of parental education among low-income families at that time.

*Head Start Impacts*

Large-scale policy lessons might be gleaned more reliably from studies of Head Start, since that program now provides services to almost a million three- and four-year-olds. Early quasi- experimental evaluations of Head Start found significant short-term gains in participants' achievement test scores, but as with Perry and Abecedarian, these achievement gains appeared to fade over time (Cicirelli 1969; McKey et al. 1985). Despite methodological critiques of these early studies (McGroder 1990), a random-assignment national study of Head Start was not undertaken for another 30 years.

The Head Start Impact Study (HSIS) used waitlist lotteries to assign children to the opportunity to enroll in a Head Start program. Begun in 2002, results indicated that after one academic year in the program, four-year olds who had the opportunity to enroll in Head Start gained significantly more in six language and literacy areas than control-group children who lost the enrollment lotteries, with these intent-to-treat effect sizes ranging from .09 to .31 standard deviations (U.S. Department of Health and Human Services 2005). In contrast, there were few program impacts on math skills or on children's attention, anti-social, or mental health problems. The vast majority of early childhood education evaluations do not track families who declined program services, and thus provide only estimates of impacts on treated children. However, in the case of the Head Start Impact Study there were children assigned to the treatment who did not experience Head Start, treatment-on-the-treated estimates take this "non-compliance" with treatment status into account, and resulting effect sizes were roughly 50 percent larger than intent-to-treat effect sizes (Ludwig and Phillips 2007). By the end of first grade, both achievement levels and behavioral ratings of treatment group children were essentially similar to achievement levels of control-group children (U.S. Department of Health and Human Services 2010).

Why might Head Start's initial achievement impacts disappear so quickly? All children learn, but they learn at different rates. If the test scores of Head Start and comparison-group children converge during elementary school, then the treatment group's preschool gains must be offset later by larger gains in the control group. Why this happens is not entirely clear; most arguments focus on the quality of subsequent schools that children attend. If little learning occurs in low-quality schools, then early advantages imparted by programs such as Head Start might be lost. In this case, preschool does not "immunize" against the adverse effects of subsequent low-quality schooling (Currie and Thomas 2000; Lee and Loeb 1995).

As part of their examination of why Head Start impacts fade out more rapidly for African-American children than for whites, Currie and Thomas (2000) show that African-American children in Head Start attend lower-quality schools, as measured by students' average test scores, relative to the schools attended by African-American children who did not attend Head Start. For white children, in contrast, average school quality did not differ by Head Start participation status. Similarly, Zhai, Raver, and Jones (2012) find that the benefits to children of an intervention designed to enhance the developmental quality of Head Start programs persisted into kindergarten only for those children who attended relatively higher-quality elementary schools, again measured by student test scores.

An alternative explanation of achievement-impact fadeout is that kindergarten teachers might be particularly effective at teaching children with low levels of skills. In this case, it may be that the classroom is not of generally low quality, but instructional efforts may favor children at the lower end of the skill distribution, which would include larger concentrations of children who had not participated in early childhood education. Indirect evidence supporting this hypothesis is provided in the work of Engel, Claessens, and Finch (2011), who find that kindergarten teachers spend the most time on very basic math instruction (like learning numbers) despite the fact that the vast majority of kindergarteners have already acquired such skills. If this explanation holds, the effects of early childhood education programs are most likely to persist in subsequent schooling environments in which learning gains are equally distributed across children with high and low levels of initial skills.[3]

As with Perry and Abecedarian, quickly declining test score impacts for recent cohorts of Head Start children appear to be at odds with the long-term impacts on important young adult outcomes found in analyses of older Head Start cohorts. Some of the older-cohort studies use strong quasi-experimental methods and find quite striking long-run program impacts. One of the most recent and comprehensive is Deming's (2009) sibling-based fixed-effect analysis, which found that compared with siblings who did not attend Head Start or other preschool programs, children who attended Head Start in the 1980s and early 1990s were over 8 percentage points more likely to graduate from high school.  Deming's more general composite of positive early adult outcomes—including high school graduation, college attendance, idleness, crime, teen parenthood, and health status—shows an estimated impact of .23 standard deviations.

Ludwig and Miller's (2007) regression discontinuity study of Head Start attendees in the late 1960s found that successful efforts to increase the likelihood that poor counties would establish Head Start programs by providing federal grant-writing assistance led to 3-4 percentage points gains in high school graduation rates and post-secondary schooling in the 1990 census data relative to counties with very similar levels of poverty that were not offered such assistance, although such effects were attenuated by 2000.  Taken together, these studies suggest that despite the decline in program impacts on achievement test scores as children progress through elementary school, there may be measureable and important, effects of Head Start on children's life chances.

*Pre-Kindergarten Programs*

Some rigorous evaluations of pre-kindergarten programs were completed too recently to have been included in the database used to produce Figure 2. Most of these studies use regression discontinuity designs based on strict birthday cutoffs, and thus compare past and prospective pre-kindergarten program attendees rather than entire cohorts of children. For this reason (and a few others), these designs are not directly comparable to either intent-to-treat or treatment-on-the-treated estimates from experimental studies (Weiland, Yoshikawa, Wilson, Hofer, Dong, and Lipsey 2011; Gibbs, Ludwig, and Miller 2012). The most comprehensive overview is Wong et al. (2008), which examines five state pre-kindergarten programs and finds short-run effects on achievement test scores that are somewhat larger than those estimated in the National Head Start Impact Study, although the size of the impacts varies considerably across states and types of test (weighted average intent-to-treat impacts range from .17 for vocabulary to .68 for print awareness).

The highly regarded Tulsa pre-kindergarten program has also been carefully evaluated. A birthday cutoff-based regression discontinuity evaluation of the program found large and significant effects on children's achievement, with effect sizes ranging from .38 to .79 (Gormley, Gayer, Phillips, and Dawson 2005). Using propensity score matching methods to adjust for differences in children's backgrounds, the researchers found that the Tulsa pre-kindergarten program reduced attendees' timidity and improved their attentiveness. The program did not appear to affect disobedience, apathy, aggression, learning task problems, or problems interacting with peers or teachers (Gormley, Phillips, Newmark, Welti, and Adelstein 2011).

The only longer-run follow-up study conducted to date of pre-kindergarten program uses propensity matching and administrative data on third grade test scores. Estimating program impacts for two cohorts, Hill and colleagues (2012) find no lasting discernible achievement impacts for the first cohort by third grade, and for the second cohort there is evidence of persisting math impacts (.18 standard deviations), perhaps reflecting an increased emphasis on math instruction, including the introduction of new curricula, in during elementary school. The lack of longer-run evaluations of pre-kindergarten programs suggests that drawing strong policy conclusions about their effectiveness is unwarranted, as other programs have likewise demonstrated early promising results that faded over the first few years of school.

**The Puzzle: Academic Fade-Out but Long-Term Benefits**

Most early childhood education studies that have tracked children beyond the end of the program treatment find that effects on test scores fade over time. An analysis of cognitive and achievement outcomes in our meta-analytic database, which includes model programs such as Perry Preschool as well as Head Start and many other programs, shows an estimated decrease in program impact effect sizes of about .03 standard deviations per year. With end-of-treatment effect sizes averaging around .30 standard deviations, this implies that positive effects persist for roughly 10 years (Leak et al. 2011; see also Aos et al. 2004; Camilli et al. 2010). This finding raises a puzzle: How do we reconcile the fade-out of preschool program impacts on test scores during elementary school with the evidence showing that such programs nonetheless have beneficial impacts on a broad set of later-life outcomes like high school graduation rates, teen parenthood, and criminality?

One obvious possible explanation is that preschool programs may affect something other than basic achievement and cognitive test scores, and perhaps these other program impacts, unlike achievement and cognitive impacts, persist over time. In turn, this raises the question of exactly how early childhood education programs affect various aspects of development, including cognitive skills, personality traits like conscientiousness, and the behavior categories like attentiveness or antisocial behavior that are often emphasized by development psychologists? The literature on the effects of preschool has drawn on several different models of human development.

In one prominent example, Cunha and Heckman (2007) posit a cumulative model of the production of human capital that allows for the possibility of differing childhood investment stages as well as roles for the past effects and future development of both cognitive and socio-emotional skills. In this model, children have endowments at birth of cognitive potential and temperament that reflect a combination of genetic and prenatal environmental influences. The Cunha and Heckman model highlights the interactive nature of skill building and investments from families, preschools and schools, and other agents. It suggests that human capital accumulation results from "self-productivity" – skills developed in earlier stages bolster the development of skills in later stages – as well as the dynamic complementary that results when skills acquired prior to a given investment increase the productivity of that investment. These two principles are combined in the hypothesis that "skill begets skill."

Several aspects of this model are relevant for preschool investment policy. If focused on the preschool period, the Cunha and Heckman (2007) model implies that school readiness is a product of the child's cognitive and socioemotional skills upon entry into the preschool period, plus preschool-period investments from parents and possibly from an early childhood education program. The hypothesis of dynamic complementarity implies that the effects of parental and early childhood education investments on child outcomes will be largest for children who enter the preschool period with the highest levels of cognitive and socioemotional skills.

Different predictions emerge from models of human capital development proposed in the developmental psychology literature. These models, too, focus on how individuals' endowments interact with environmental experiences, and suggest that both individual capacities and experience shape development (Blair and Raver 2012). However, they diverge from the Cunha and Heckman (2007) model by differentiating among types of investments, and by distinguishing how environments and investments interact to shape development. Specifically, "compensatory" models posit that preschool investments can function as a substitute for enriched home environments (Ramey and Ramey 1998). Thus, children whose skill development may be compromised by economic disadvantage or low-quality home environments are predicted to benefit more from early childhood education programs than more advantaged children. This hypothesis provided the rationale for the initial and continued funding for programs such as Head Start and Early Head Start, which target children from disadvantaged backgrounds.

If early childhood education programs seek to build children's early skills to generate lasting changes in adults' human capital, which skills should they target? Economists tend to lump IQ and achievement into a "cognitive" category and everything else into a "noncognitive" category, but this distinction is unhelpful for a variety of reasons. A first problem is that "cognitive" skills are a heterogeneous mixture of "achievement" and more basic cognitive capacities. Although scores on tests of cognitive ability and achievement tend to be highly correlated, there is an important conceptual difference between them. "Achievement" commonly refers to concrete academic skills such as literacy and numeracy that develop in response to parenting, schooling, and other human capital investments, including early childhood education, whereas IQ or general cognitive ability is considered to be a relatively more stable trait.

Second, learning skills such as the ability to sustain attention when performing tasks, plan ahead and control emotions in the face of provocation involve many of the same elements of brain circuitry as learning concrete skills, and are therefore inherently "cognitive." Third and most important, conceptualizing and measuring distinct components of "noncognitive" skills is a vital first step in understanding why early childhood education and other human capital inventions have an effect. Different branches of psychology typically categorized non-cognitive skills in very different ways.

Most personality psychologists have centered their work on the "big five" personality traits, which are derived from factor analyses of observer and self-reports of behaviors and include conscientiousness, openness, agreeableness, emotional stability, and extraversion – plus general cognitive ability.  Education research consistently shows that conscientiousness best correlates with overall attainment and achievement (Almlund, Duckworth, Heckman, and Kautz 2011). Although these traits have traditionally been viewed as relatively stable across the life span, some evidence indicates that that they can change in response to life experiences and interventions (for example, Roberts, Walton, and Viechtbauer 2006; Almlund et al. 2011).

Developmental psychologists view children's skills and behaviors as determined by the interplay between their innate abilities, dispositions and the quality of their early experiences – which may include early childhood education (National Research Council 2000). They classify skills and behaviors in a number of ways, and some of their categories corrrespond to the "big five" personality traits. For example, our own recent review classified important competencies into four groups: achievement, attention, externalizing behavior problems, and mental health (Duncan and Magnuson 2011). Attention refers to the ability to control impulses and focus on tasks (for example, Raver 2004). Externalizing behavior refers to a cluster of related behaviors including antisocial behavior, conduct disorders, and more general aggression (Campbell, Shaw, and Gilliom 2000). Mental health constructs include anxiety and depression as well as somatic complaints and withdrawn behavior (Bongers et al. 2003). All of these skills and behaviors might respond to investments in early childhood education.

Testing and comparing how these theories of human development apply in the context of early childhood education is difficult, because despite arguments that early childhood education programs are likely to generate broad impacts on children's behavior and social competence (Zigler and Trickett 1978), most preschool studies do not measure many of these kinds of outcomes at program completion. Some studies have included measures of problem behavior, typically ratings of children's antisocial or aggressive behaviors, with mixed results. Perry significantly reduced problem behavior, especially among boys, and the examination by Heckman, Moon, Pinto, Savelyev, and Yavitz (2010) of Perry's long-run effects finds that these behavior impacts explain a substantial proportion of the program's effects on boys' crime and employment outcomes. However, both early cognitive and behavioral impacts explain program impacts on girls' later outcomes. Moreover, for both genders a substantial share of the program impacts on adult outcomes is not explained by any of the observed early program impacts.

Other programs provide little evidence of program impacts on children's behavior. Deming's (2009) analysis of Head Start found no short-run effects of Head Start on parental reports of children's behavior problems. Haskins (1985) reported that the Abecedarian program had the unexpected effect of increasing teacher reports of children's aggressiveness in the early school years, although these effects appeared to fade with time. Of course, these studies are vulnerable to the criticism that they did not measure a broader set of relevant skills, including student's attention or other aspects of their behavior and mental health.

Overall, reconciling disparate patterns of impacts in the short and longer term is a key challenge for anyone hoping to extract policy lessons about the effectiveness of early childhood education programs. Accomplishing this task will require a proven model of human development that incorporates various cognitive, personality, and behavioral dimensions and can predict what kinds of children stand to benefit most from early childhood education investments.

## Within-Program Heterogeneity

Although policymakers appropriately care most about the average impacts of early childhood education programs, a number of lessons can be learned from looking at the distribution of treatment effects of given programs. For example, such heterogeneity might make it possible to identify groups that could particularly benefit from preschool programs. Data on treatment-effect heterogeneity may also boost our understanding of human capital development processes, if they identify groups that particularly benefit from the preschool setting.

Consider evidence from the Infant Health and Development Program (IHDP), shown in Figure 3. Beginning shortly after a child's birth, the IHDP offered a package of services that included a full-day, cognitively enriching curriculum for children between ages one and three, modeled after the Abecedarian program. Nearly 1,000 children in eight sites across the country were randomly assigned to the IHDP treatment or to a control group that received no early childhood education services, but some health services (Gross, Spiker, and Haynes 1997). To be eligible for the program, infants had to have weighed less than 2,500 grams (5.5 pounds) at birth, but eligibility was not restricted by family income, race, or ethnicity.

For the economically disadvantaged children in the sample—those with family income below 180 percent of the poverty line in their first year of life—participation in IHDP produced large impacts on cognitive development. Specifically, children in the treatment group outscored their control-group counterparts by .82 standard deviations on the Stanford-Binet IQ mental subscale by age three.[4] For children in higher-income families, the IHDP's program impact was much smaller – .18 standard deviations. Thus, if "disadvantage" is defined by family income, IHDP treatment impacts heavily favored disadvantaged infants. However, an alternative definition of "disadvantage" can lead to a different conclusion. Children disadvantaged by being born with a "very low" birth weight (less than 1500 grams or 3.3 pounds) benefited significantly less from the IHDP intervention than "advantaged" heavier babies in this low birth weight sample.

It is not difficult to generate possible explanations for these patterns. For example, the income results are consistent with theories positing that the IHDP focus on enriched early learning compensates or substitutes for lower levels of parental investment and academic stimulation in low-income families. The differences by birth weight are consistent with the "skill begets skill” perspective. Potential gains for very low birth weight babies’ cognitive development may be constrained by neurological challenges that the program was unable to address. In other words, the match between what the program provided and children’s individual differences may explain why some disadvantaged groups show larger impacts, but not others.

A systematic accounting of heterogeneity in the effects of preschool programs is a complicated undertaking. For example, Anderson’s (2008) study of three researcher-designed early childhood education programs—Perry, Abecedarian and the Early Training Project— showed much larger benefits for girls than boys. Turning to our meta-analytic database, we found 22 programs that estimated program impacts by gender. Anderson’s conclusions were supported by the published evaluations of his three studies, but not by the larger set of evaluation studies, as shown in Figure 4 (Kelchen et al. 2012). The first comparison shows that on cognitive and achievement outcomes, the effect is slightly larger for females. However, the second comparison shows that when a broad set of school outcomes are considered, including special education, grade retention, and other aspects of general school adjustment, boys appear to benefit much *more* from these programs than girls. Looking just at the three programs in Anderson (2008), the “other school outcomes” variable strongly favors females, so this difference in findings is generated by the inclusion of results from a broader set of studies. For the adult outcomes, females are favored, but the difference is not significant.

Even when studies determine that a particular program has been a success overall, on average, the positive outcomes differ across programs and populations. For example, Perry Preschool and Head Start significantly reduced criminal activity, but Abecedarian did not. Garces, Thomas, and Currie (2002) found that Head Start increased educational attainment for whites, but not for blacks, and led to reductions in crime for blacks but not whites.

There is much more to be learned about heterogeneity in the effects of preschool programs, although efforts to identify differential effects can be hampered by small sample sizes and limited baseline information, especially in the older studies. The program and population specificity of program impacts argues against a single explanation for how preschool programs improve long-run outcomes. Greater attention should be given to understanding both who benefits the most from particular programs and why.

## The Search for Active Program Ingredients

Research on early childhood education has focused greater attention on evaluating particular programs than on identifying the particular ingredients in these programs that produce significant improvements in children's learning and behavior. The research problem here is difficult. For example, some scholars have focused on structural aspects of early childhood education environments, such as class size and teacher education, yet these features of programs are likely to affect children only indirectly, by influencing their experiences within classrooms. Perhaps not surprisingly, associations between these features of classrooms and preschoolers' learning are inconsistent and weak (Mashburn et al. 2008).[5]

Much harder to measure than class size or teacher education, but potentially more important for children's actual experiences in early childhood education programs, is what developmental psychologists have referred to as "process quality" -- the quality of classroom interactions, including the amount of instructional and emotional support children receive. Associations between these aspects of process quality and children's outcomes are more consistently positive, if still modest (Burchinal, Kainz, and Cai, in press). As attention has shifted to improving classroom interactions, two aspects of program design emerge as policy levers that may, together, improve program effectiveness: curriculum and related professional development. To cite one example, best practices for mathematics instruction explicitly incorporate foundational math conceptual learning within everyday activities and provide activities that support a developmental progression of mathematical learning (Clements and Samara 2011). Despite the identification of best practices and the availability of curricula that provide lesson plans, research consistently finds that the instructional quality of most preschool classrooms is poor (Justice, Mashburn, Hamre, and Pianta 2008).

It appears that an effective strategy is to combine a proven curriculum that offers well-designed lesson plans and activities based on an understanding of children's trajectories of learning within specific content areas, with strong professional development to target improvement in specific instructional practices. Several random-assignment studies of curricular innovations in early childhood education programs have shown substantial effects on children's learning in math and literacy, and these curricula are currently found in some effective preschool programs. The What Works Clearinghouse provides up-to-date information on rigorous evaluations of early childhood education curricula (http://ies.ed.gov/ncee/wwc/).

The Boston pre-kindergarten system provides a scaled-up model of how this might work. System leaders developed a curriculum from proven literacy, math and social skills interventions. The academic components focused on concept development, the use of multiple methods and materials to promote children's learning, and a variety of activities to encourage analysis, reasoning, and problem-solving (Weiland and Yoshikawa, forthcoming). Extensive professional development training and on-going coaching ensure that teachers understood the curriculum and were able to implement it effectively in their classrooms. A regression-discontinuity evaluation showed relatively large impacts on vocabulary, math, and reading (effect sizes ranging from .45 to .62 standard deviations) as well as smaller, but still noteworthy effects on working memory and inhibitory control (effect sizes ranging from .21 to .28 standard deviations; Weiland and Yoshikawa 2012).

**Conclusions**

Theories and evidence across the social sciences argue that early childhood may be a promising period for effective educational investments, particularly for disadvantaged children. Early cognitive and socioemotional skills are sensitive to environmental inputs, and building skills early in life may produce lasting effects. Most evaluations of early education programs show that such programs improve children's school readiness, specifically their pre-academic skills, although the distribution of impact estimates is extremely wide, and gains on achievement tests typically fade over time. Some studies of children who attended preschool 20 or more years ago find that early childhood education programs also have lasting effects on children's later life chances, improving educational attainment and earnings and, in some cases, reducing criminal activity. High-quality early childhood education programs thus have the potential to generate benefits well in excess of costs. Despite general agreement about these aspects of early childhood education studies, important questions about the wisdom of large-scale investments in early childhood education remain unanswered.

First, we need to know much more about how early childhood education works: that is, connections between program components and particular child outcomes. Since program impacts on cognitive ability and achievement often fade within a few years of the end of the programs, these skills do not appear to be driving longer-run effects. Data constraints have made it difficult to identify the other skills, behaviors, or developmental processes that lead to such positive outcomes in early adulthood, but efforts to better identify and measure likely pathways are critical for improving our understanding of human capital accumulation and judging whether policy and programmatic efforts are worthwhile investments. It also important to think about what programs (or parts of programs) might be scaled up in a cost-effective manner.

Second, we need a better understanding of the pattern of these program effects over time. This is likely to require new data collection efforts, because administrative data about participation in these programs, demographic background, and scores on various tests are unlikely to provide necessary information on the full range of attention, behavior, and mental health measures.

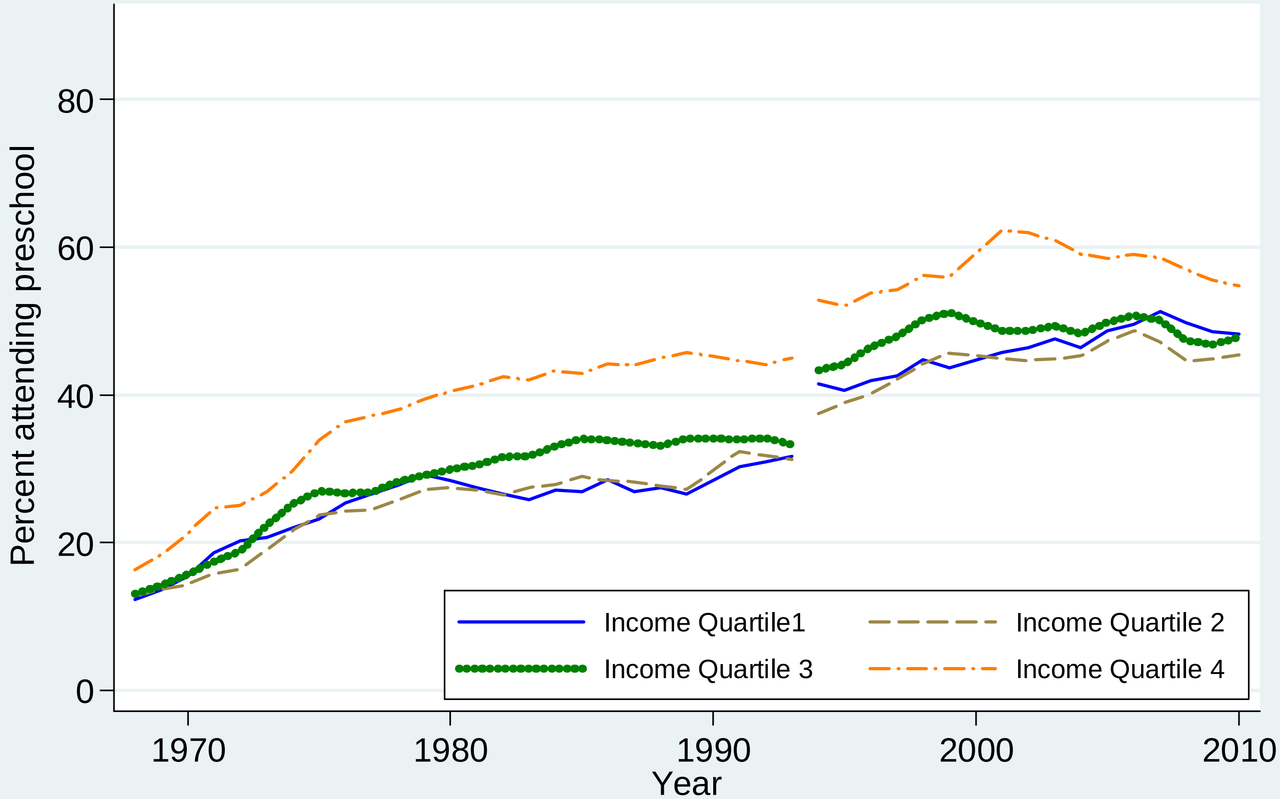
Finally, we need a more complete understanding of which skills, or constellation of skills, are likely to produce improved outcomes later in life. This requires not only understanding how programs affect later skills, but also a better grasp of how skills, behavior, attention, and mental health in childhood build human capital and other labor market outcomes in adulthood.

Given the potential payoff from early education and the importance of early skills in forecasting later school and labor market success, supporting low-income children's participation in high-quality early childhood investment may well constitute a wise investment.  The potential for profitable investments exists at both margins – enrolling low-income children who are not currently attending a preschool program as well as improving the quality of existing programs, although we know more about the former than the latter (Duncan, Ludwig and Magnuson 2010). What may be more important in the long term than any specific programmatic change is a change in how research is conducted in this area. Rather than looking merely at average short-run outcomes of early childhood education programs based on a limited number of achievement tests, researchers should focus on the heterogeneity of outcomes across groups, conduct long-term follow-up, and examine a wide range of outcome variables that would illuminate the program ingredients and developmental processes that make some of these programs so successful.

**Acknowledgment**

This paper draws extensively from our collaborations with Hirokazu Yoshikawa, Holly Schindler and Jack Shonkoff. We are grateful to the following funders of the National Forum on Early Childhood Policy and Programs: the Birth to Five Policy Alliance, the Buffett Early Childhood Fund, Casey Family Programs, the McCormick Tribune Foundation, the Norlien Foundation, Harvard University, and an anonymous donor. We would also like to thank the Institute of Education Sciences (R305A110035) for supporting this research, to Abt Associates, Inc. and the National Institute for Early Education Research for making their data available to us, and to Weilin Li for helpful research assistance.

Figure 1: Percent of children 3- and 4-year-olds enrolled in preschool
by family income quartile




Note data are taken from the October CPS, and represent 3 year moving averages. Parents report on whether the child attends "regular school." The line break in 1994 corresponds to the addition of a question prompt, which defined regular school as including “nursery school, kindergarten or elementary school...". See Magnuson et al. (2007) for further discussion of how the CPS compares with other sources of data on preschool enrollment.

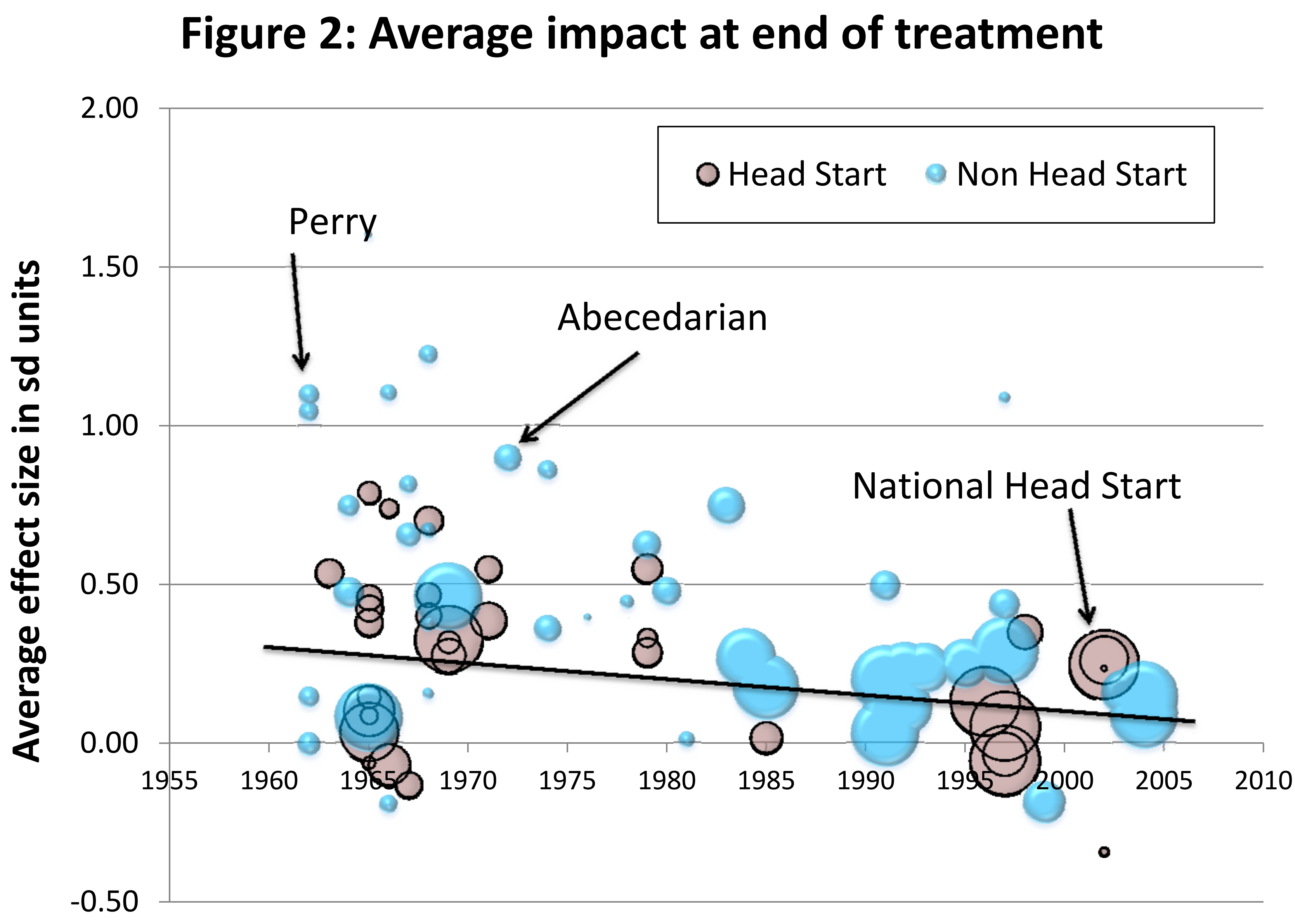
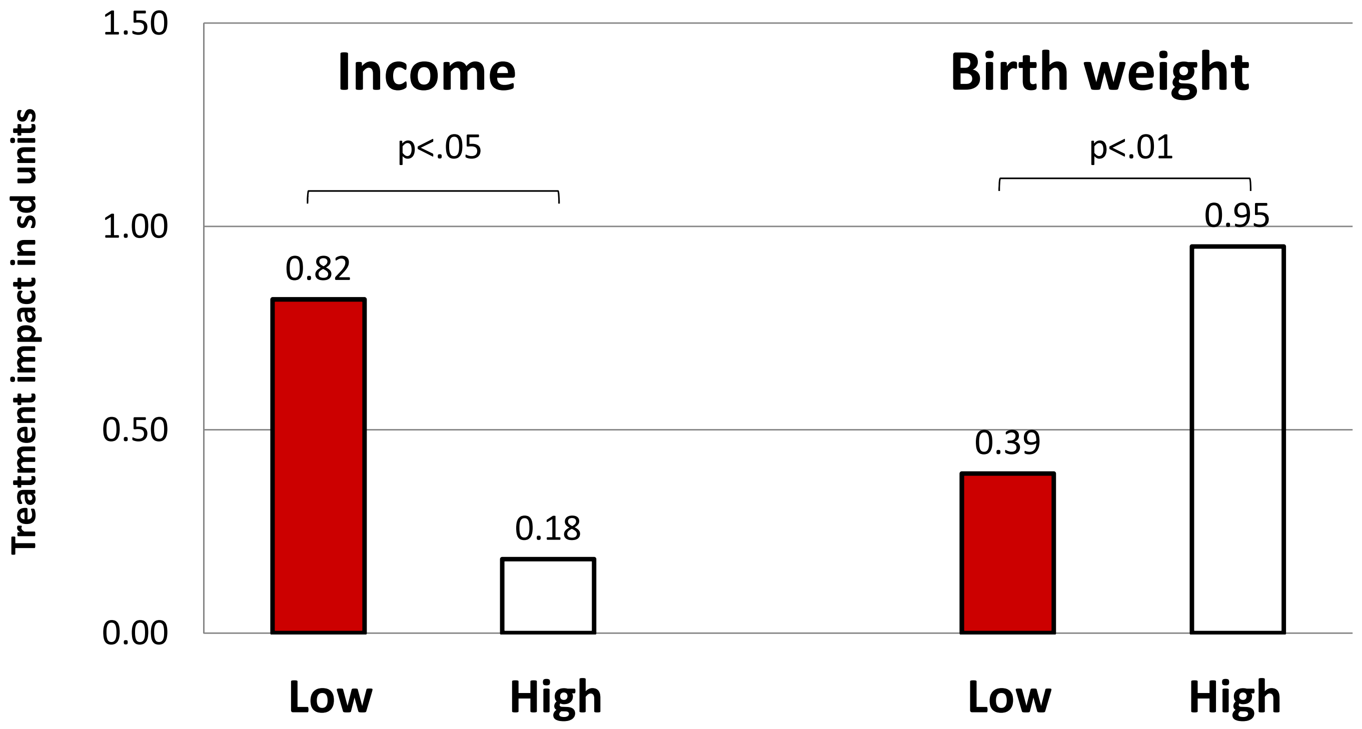


Figure 2: Average impact at end of treatment
Head Start
Non Head Start
Perry
Abecedarian
National Head Start
Average effect size in sd units
2.00
1.50
1.00
0.50
0.00
-0.50
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010

## Figure 3: Impacts of the IHDP treatment on age-3 IQ by income and birth weight




All models also condition on child gender, birth weight, gestational age at birth, neonatal health index and site indicators.

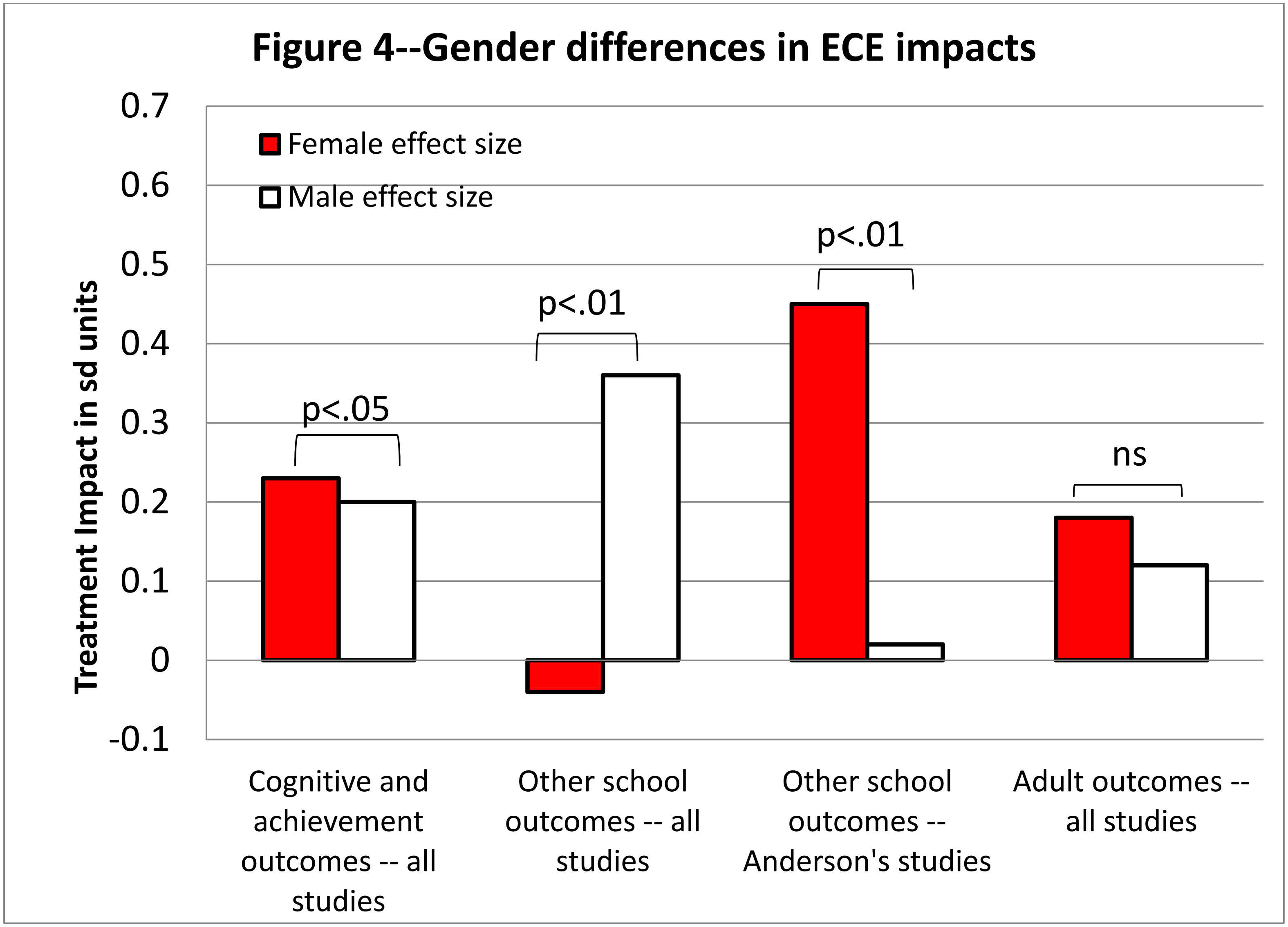


Figure 4--Gender differences in ECE impacts
Female effect size
Male effect size
Treatment Impact in sd units
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
p<.05
p<.01
p<.01
ns
Cognitive and achievement outcomes -- all studies
Other school outcomes -- all studies
Other school outcomes -- Anderson's studies
Adult outcomes -- all studies

REFERENCES


Adams, Gina, and Monica Rohacek. 2002. More Than a Work Support? : Issues Around Integrating Child Development Goals Into The Child Care Subsidy System. *Early Childhood Research Quarterly 17* (4): 418-440.

Almlund, Mathilde, Angela L. Duckworth, James J. Heckman, and Tim Kautz. 2011. Personality Psychology and Economics. IZA Discussion Paper 5500.

Anderson, Michael. 2008. Multiple Inference and Gender Differences in the Effects Of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association, 103* (484): 1481-1495.

Aos, Steve, Roxanne Lieb, Jim Mayfield, Marna Miller, and Annie Pennucci. 2004. Benefits and Costs of Prevention and Early Intervention Programs for Youth. Olympia: Washington State Institute for Public Policy.

Barnett, W. Steven, Megan Carolan, Jen Fitzgerald, and James Squires. 2011. The State of Preschool 2011: State Preschool Yearbook. National Institute for Early Education Research, Rutgers.

Barnett, W. Steven and Leonard N. Masse. 2007. Comparative Cost-Benefit Analysis of the Abecedarian program and Its Policy Implications. *Economics of Education Review 26*(1): 113-125.

Besharov, Douglas J. and  Caeli A. Higney. 2007. Head Start: Mend It, Don't Expand It (Yet). *Journal of Policy Analysis and Management, 26* (3): 678-681.

Blair, Clancy and C. Cybele Raver. 2012. Child Development in the Context of Poverty: Experiential Canalization of Brain And Behavior.  *American Psychologist 67* (4): 309-318.

Burchinal, Margaret, Kirsten Kainz, and Karen Cai. In Press. How Well Do Our Measures of Quality Predict Child Outcomes? A Meta-Analysis and Coordinated Analysis of Data from Large-Scale Studies of Early Childhood Settings. In M. Zaslow (Ed.) *Reasons to take stock and strengthen our measures of quality*. Baltimore, MD: Brooks Publishing.

Bongers, Ilja L., Hans M. Koot, Jan van der Ende, and Frank C. Verhulst. 2003. The Normative Development of Child and Adolescent Problem Behavior. *Journal of Abnormal Psychology 112* (5): 179–92.

Camilli, Gregory, Sadako Vargas, Sharon Ryan, and W. Steven Barnett. 2010. Meta-analysis of the Effects of Early Education Interventions on Cognitive and Social Development. *Teachers College Record, 112* (3): 579-620.

Campbell, Susan B., Daniel S. Shaw, and Miles Gilliom. 2000. Early Externalizing Behavior Problems: Toddlers and Preschoolers at Risk for Later Adjustment. *Development and Psychopathology* 12(3): 467–88.

Campbell, Frances A., Craig T. Ramey, Elizabeth Pungello, Joseph Sparkling, and Shari Miller-Johnson. 2002. Early Childhood Education: Young Adult Outcomes from the Abecedarian Project. *Applied Developmental Science 6* (1): 42-57.

Chetty, Raj, John Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Schanzenbach, and Danny Yagan. 2011. How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR, *Quarterly Journal of Economics* 126(4): 1593-1660.

Card, David E. 1999. The Causal Effect of Education on Earnings. In O. Ashenfelter, & D. Card (Eds.), *Handbook of Labor Economics* (Vol. 3, pp. 1801–63). Amsterdam: North-Holland.

Cicirelli, Victor G. 1969. *The Impact of Head Start: An Evaluation of the Effects of Head Start on Children's Cognitive and Affective Development*. Report presented to the Office of Economic Opportunity pursuant to contract B89-4536 (report No. PB 184 328). Westinghouse Learning Corporation. Washington, DC: National Bureau of Standards, Institute for Applied Technology**.**

Clements, Douglas and Julie Samara. 2011. Early Childhood Mathematics Intervention. *Science 333*: 968-970.

Cunha, Flavio, and James J. Heckman. 2007. The Technology of Skill Formation. *American Economic Review 97* (2): 31–47.

Currie, Janet. 2001. Early Childhood Education Programs. *Journal of Economic Perspectives 15* (2): 213-238.

Currie, Janet, and Duncan Thomas. 1995. Does Head Start Make a Difference? *American Economic Review 85* (3): 341-364.

Currie, Janet, and Duncan Thomas. 2000. School Quality and the Longer-Term Effects of Head Start. *Journal of Human Resources 35* (4): 755-774.

Deming, David. 2009. Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start. *American Economic Journal: Applied Economics, 1*(3): 111-34.

Duncan, Greg, and Aaron Sojourner. forthcoming. Can Intensive Early Childhood Intervention Programs Eliminate Income-Based Cognitive and Achievement Gaps? *Journal of Human Resources.*

Duncan, Greg, and Katherine Magnuson. 2011. The Nature and Impact of Early Achievement Skills, Attention and Behavior Problems. In Greg J. Duncan and Richard J. Murnane (eds.), *Whither Opportunity: Rising Inequality, Schools, and Children's Life Chances*, New York: Russell Sage, pp. 47-69.

Engel, Mimi, Amy Claessens, and Maida Finch. 2012. Teaching Students What They Already Know? The Misalignment between Instructional Content in Mathematics and Student Knowledge in Kindergarten. Working paper, Peabody College, Vanderbilt University.

Garces, Eliana, Duncan Thomas, and Janet Currie. 2002. Longer Term Effects of Head Start. *The American Economic Review 92:* 999-1012.

Gibbs, Chloe, Jens Ludwig, and Doug Miller, D. 2012. Does Head Start Do Any Lasting Good? NBER Working Paper 17452.

Gormley, William T., Deborah Phillips, Katie Newmark, Kate Welti, and Shirley Adelstein. 2011.Social-Emotional Effects of Early Childhood Education Programs in Tulsa, *Child Development 82*(6): 2095-2109

Gormley, William T., Jr., Ted Gayer, Deborah Phillips, and Brittany Dawson. 2005. The Effects of Universal Pre-K on Cognitive Development. *Developmental Psychology, 41*(6): 872-884.

Gross, Ruth T., Donna Spiker, and Christine Haynes. 1997. Helping Low Birth Weight, Premature Babies: The Infant Health and Development Program. Stanford, CA: Stanford University Press.

Ha, Yoonsook, Katherine Magnuson and Marcy Ybarra. 2012. The Association between Child Care Subsidies and Stability of Care. *Children and Youth Services Review* 34, 1834-1844.

Hamm, Katie. 2006. *More than Meets the Eye: Head Start Programs, Participants, Families, and Staff in 2005.* Head Start Series, Brief No. 8. Washington, DC: Center for Law and Social Policy.

Haskins, Ron. 1985. Public School Aggression Among Children with Varying Day-Care Experience, *Child Development 56* (3): 689-703.

Heckman, James J. 2006. Skill Formation and the Economics of Investing in Disadvantaged Children *Science* 30 June 2006: 312 (5782), 1900-1902. [DOI:10.1126/science.1128898]

Heckman, James J., Seong Hyeok Moon, Rodrigo Pinto, Peter Savelyev, Adam Yavitz. 2010. A New Cost-Benefit and Rate of Return Analysis for the Perry Preschool Program: A Summary, NBER Working Paper 16180.

Hill, Carolyn, William Gormley, and Shirley Adelstein. 2012. Do the Short-Term Effects of a Strong Preschool Program Persist? Georgetown University Center for Research on Children in the United States Working Paper # 18.

Justice, Laura M., Andrew Mashburn, Bridget K. Hamre, and Robert C. Pianta. 2008. Quality of language and literacy instruction in preschool classrooms serving at-risk pupils. *Early Childhood Research Quarterly 23(1)*: 51-68.

Kelchen, Robert, Katherine Magnuson, Greg Duncan, Holly Schindler, Hilary Shager, and Hirokazu Yoshikawa. 2012. Do the Effects of Early Childhood Programs on Academic and Adult Outcomes Vary by Gender? A Meta-Analysis. Working Paper.

Knudsen, Eric, James J. Heckman, Judy Cameron, and Jack Shonkoff. 2006. Economic, Neurobiological and Behavioral Perspectives on Building America's Future Workforce. *Proceedings of the National Academy of Sciences of the United States of America, 103* (27): 10155-62.

Leak, James, Greg Duncan, Weilin Li, Katherine Magnuson, Holly Schindler, and Hirokazu Yoshikawa. 2012. Is Timing Everything? How Early Childhood Education Program

Cognitive and Achievement Impacts Vary by Starting Age, Program Duration and Time since the End of the Program. Working Paper.

Lee, Valerie and Susanna Loeb. 1995. Where Do Head Start Attendees End Up? One Reason Why Preschool Effects Fade Out. *Educational Evaluation and Policy Analysis, 17*(1): 62-82.

Love, John M., Linda Harrison, Abraham Sagi-Schwartz, Marinus van IJzendoorn, Christine Ross, Judy Ungerer. et al. 2003. Child Care Quality Matters: How Conclusions May Vary With Context. *Child Development 74* (4): 1021–1033.

Ludwig, Jens, and Douglas L. Miller. 2007. Does Head Start Improve Children's Life Chances: Evidence from a Regression Discontinuity Design. *Quarterly Journal of Economics, 122* (1): 159-208.

Ludwig, Jens, and Deborah Phillips. 2007. The Benefits and Costs of Head Start. *Social Policy Report, 21* (3): 3-20.

Mashburn, Andrew, Robert Pianta, Bridget Hamre, Jason Downer, Oscar Barbarin, Donna Bryant, Margaret Burchinal, Diane Early, and Carolee Howes. 2008. Measures of Classroom Quality In Prekindergarten and Children's Development of Academic, Language, and Social Skills. *Child Development, 79* (3): 732-749.

Magnuson, Katherine, Claudia Lahaie, and Jane Waldfogel. 2006. Preschool and School Readiness of Children of Immigrants.  *Social Science Quarterly 87*: 1241-1262.

Magnuson, Katherine, Marcia Meyers,and Jane Waldfogel. 2007. Public Funding and Enrollment in Formal Child Care in the 1990s. *Social Service Review, 81* (1): 47-83.

Magnuson, Katherine, and Hilary Shager. 2010. Early Education: Progress and Promise for Low-income Children. *Children and Youth Services Review, 32*: 1186 – 1198

McGroder, Sharon. 1990. *Head Start: What Do We Know About What Works?* Report prepared for the U. S. Office of Management and Budget. Retrieved June 30, 2012, from http://aspe.hhs.gov/daltcp/reports/headstar.pdf.

McKey, Ruth  Hubbell, Larry Condelli, Harriet Ganson, Barbara J. Barrett, B. J., Catherine C. McConkey, C., and Margaret C. Plantz, M. C. 1985. *The Impact of Head Start on Children, Families and Communities: Final Report of the Head Start Evaluation, Synthesis and Utilization Project.* Washington, D. C.: CSR, Incorporated.

National Association of Child Care Resource & Referral Agencies. 2011. *Parents and the High Cost of Child Care.* Arlington, VA: NACCRRA.

National Institute of Child Health and Human Development Early Childcare Research Network, and Greg J. Duncan. 2003. Modeling the Impacts of Child Care Quality on Children's Preschool Cognitive Development. *Child Development, 74* (5): 1454-1475.

National Research Council. 2000. *From Neurons to Neighborhoods: The Science of Early Childhood Development*. Shonkoff, Jack, and Deborah Phillips, eds. Washington, D.C.: National Academy Press.

Nelson, C. A., & Sheridan, M. A. (2011). Lessons from neuroscience research for understanding causal links between family and neighborhood characteristics and educational outcomes. Pp. 27-46 in G. Duncan & R. Murnane (Eds.), *Whither opportunity: Rising inequality, schools, and children's life chances* (pp. 27-46). New York: Russell Sage.

Olds, David L., Lois Sadler, and Harriet Kitzman. 2007. Programs for Infants and Toddlers: Recent Evidence from Randomized Trials. *Journal of Child Psychology and Psychiatry, 48* (3-4): 355-391.

Ramey, Craig T., and Frances A. Campbell. 1979. Compensatory Education for Disadvantaged Children. *The School Review 87* (2): 171-189.

Ramey, Craig T., Frances A. Campbell, Margaret Burchinal, Martie Skinner, David Gardner, and Sharon Ramey. 2000. Persistent Effects of Early Childhood Education on High-Risk Children and Their Mothers. *Applied Developmental Science 4* (1): 2-14.

Ramey, Craig T., and Francis Campbell. 1984. Preventive Education for High-Risk Children: Cognitive Consequences of the Carolina Abecedarian Project. *American Journal of Mental Deficiency 88* (5): 515-523.

Ramey, Craig T., and Sharon Landesman Ramey. 1998. Early Intervention and Early Experience. *American Psychologist 53*: 109-120.

Raver, C. Cybele. 2004. Placing Emotional Self-Regulation in Sociocultural and Socioeconomic Contexts." *Child Development 75* (2): 346–53.

Roberts, Brent, Kate Walton, and Wolfgang Viechtbauer, 2006. Patterns of Mean-Level Change in Personality Traits across The Life Course: A Meta-Analysis Of Longitudinal Studies. *Psychological Bulletin 132* (1): 1-25.

Sapolsky, Robert. 2004. Mothering Style and Methylation. *Nature Neuroscience 7* (8): 791-792.

Schweinhart, Lawrence J., Jeanne Montie, Zongping Xiang, W. Steven Barnett, Clive R. Belfield, Milagros Nores. 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 Press.

Schulman, Karen and Helen Black. 2012. Downward Slide: State Child Care Assistance Policies 2012. National Women's Law Center. (retrieved from http://www.nwlc.org/sites/default/files/pdfs/NWLC2012_StateChildCareAssistanceReport.pdf on 2/12/2012).

Shager, Hilary, Holly Schindler, Katherine Magnuson, Greg Duncan, Hirokazu Yoshikawa, Cassandra Hart. 2012. Can Research Design Explain Variation in Head Start Research Results? A Meta-analysis. Working Paper, University of Wisconsin-Madison.

Tamis-LeMonda, Catherine, Marc Bornstein, and Lisa Baumwell. 2001. Maternal Responsiveness and Children's Achievement of Language Milestones. *Child Development 72* (3): 748-767.

U.S. Department of Health and Human Services, Administration for Children and Families. 2005. *Head Start impact study: First year findings*. Washington, DC.

U.S. Department of Health and Human Services, Administration for Children and Families. 2010. *Head Start impact study: Final report.* Washington, DC.

Weiland, Christina, and Hirokazu Yoshikawa. 2012. The Impacts of an Urban Public Prekindergarten Program on Children's Mathematics, Language, Literacy, Executive Function, and Emotional Skills: Evidence From Boston. Working paper, Harvard Graduate School of Education.

Wong, Vivian, Thomas Cook, W. Steven Barnett, and Kuanghee Jung. 2008. An Effectiveness-Based Evaluation of Five State Pre-Kindergarten Programs. *Journal of Policy Analysis & Management 27* (1): 122-154.

Zigler, Edward, and Penelope Trickett. 1978. IQ, social competence, and evaluation of early childhood intervention programs. *American Psychologist 33* (9): 789-798.

Zhai, Fuhua , C. Cybele Raver, and Stephanie M. Jones. 2012. Quality of subsequent schools and impacts of early interventions: Evidence from a randomized controlled trial in Head Start settings. *Children and Youth Services Review 34* (5): 946-954.

Appendix

Early Childhood Education Meta-analytic data base

The meta-analytic database used for Figure 2 is the product of the National Forum on Early Childhood Policy and Programs (http://developingchild.harvard.edu/initiatives/forum/) and was compiled under the direction of the authors, Holly Schindler and Hirokazu Yoshikawa . The Forum is housed at the Center for the Developing Child at Harvard University, which is directed by Jack Shonkoff. A multi-step data collection and evaluation process was used to determine what studies would be included in this database. The first step was to conduct a comprehensive search of the literature from 1960 to 2007, when the coding project began. The Forum was able to take advantage of another meta-analytic database compiled previously by Abt Associates, Inc. and the National Institute for Early Education Research (NIEER), which included early childhood intervention studies from 1960-2003 (Camilli et al., 2010; Jacob, Creps & Boulay, 2004; Layzer, Goodson, Bernstein & Price, 2001). This yielded 624 previously coded studies.

To search for additional studies, the Forum next conducted keyword searches in ERIC, PsycINFO, EconLit, and Dissertation Abstracts databases, resulting in 9,617 documents, as any given program may produce a series of such documents. It then manually searched the websites of policy institutes (e.g., RAND, Mathematica, NIEER) and state and federal departments (e.g., U.S. Department of Health and Human Services), as well as references mentioned in collected studies and other key early childhood education reviews. This search produced another 692 documents. In sum, 10,309 documents for potential inclusion in the early childhood education portion of the database were identified.

The vast majority (91%) of the 10,309 documents were excluded because they violated at least one of the inclusion criteria. Most of the excluded documents were not actual research studies but rather commentaries or reviews. Others were eliminated because of methodological shortcomings, because they were not clear about the selection process for participants in the study, or because their interventions took place outside of the United States. The resulting database includes studies of programs or interventions for children between birth and age 5.

The Forum next developed criteria for the inclusion of studies into its meta-analytic database. In addition to the requirement that they examine an early childhood education intervention or program reported on from 1960 to 2007, studies had to have a treatment and control/comparison group, rather than simply assessing the growth of one group of children over time. Each of the groups in the study had to have included at least 10 participants and incurred less than 50% attrition. Studies were excluded if they were testing a pharmacological agent, assessed children with medical disorders or learning disabilities, or tested the effectiveness of medical procedures or health-related products.

Figure 2 focuses exclusively on treatment effects from early childhood education programs, defined as structured, center-based early childhood education classes, day care with some educational component, or center-based child care. These include full pre-school programs such as Head Start and other interventions conducted by researchers. Programs included were required to have provided services to children, their families, or staff at the program sites, and assessed program impacts on children's cognitive and achievement outcomes.

About one-third of the ECE studies used random assignment with the remainder following quasi-experimental designs such as change models, individual or family fixed effects models, regression discontinuity, difference in difference, propensity score matching, interrupted time series, instrumental variables and some other types of matching. Studies that used quasi-experimental designs must have had pre- and post-test information on the outcome or established baseline equivalence of groups on demographic characteristics determined by a joint test.

A team of nine graduate research assistants at three universities (Harvard, UC Irvine, and University of Wisconsin-Madison) were trained as coders during a 3- to 6-month process that included instruction in evaluation methods, using the coding protocol, and computing effect sizes. Trainees were paired with experienced coders in multiple rounds of practice coding. Before coding independently, research assistants also passed a reliability test comprised of randomly selected codes from a randomly selected study. In order to pass the reliability test, research assistants had to calculate 100% of the effect sizes correctly and achieve 80% agreement with a master coder for the remaining codes. In instances when research assistants were just under the threshold for effect sizes, but were reliable on the remaining codes, they underwent additional effect size training before coding independently and were subject to periodic checks during their transition. Questions about coding were resolved in weekly research team conference calls usually involving all four principal investigators, and decisions were kept in an annotated codebook so that decisions about ambiguities could be recalled when coding subsequent studies. In a few instances, codes were added and previously coded studies were adjusted accordingly to account for the new additions.

Each study's outcome measures were coded into standardized mean difference effect sizes, which were computed using the Comprehensive Meta-Analysis computer software program (Borenstein, Hedges, Higgins, & Rothstein, 2005). We used the Hedges' *g*-based definition of effect sizes, which adjusts standardized mean differences (Cohen's d) to account for bias arising from small sample sizes.

Data for Figure 2 come from the 84 programs that assessed cognitive and/or achievement outcomes around the end of their treatments and used no-program control groups. So, for example, early childhood education programs that tested a version of a classroom-based program that included home visitation vs. the classroom program that did not include home visitation are not included in the figure.

Operationally, we choose effects sizes closest to the end of treatments but defined the end of treatment interval to include effect sizes measured as late as one year following the completion of the program and as early as 75% of the length of the program. All available effect sizes for cognitive or achievement measures at the point closest to the end of treatment are included in the analysis.

The effect sizes shown in Figure 2 and in the Appendix Table are simple means of all effect size estimates for cognitive and achievement outcomes for each given program. Bubbles areas are proportional to the inverse of squared standard errors of the average estimates, which is calculated by a Bayesian shrinkage model to take sampling variation of the within-study estimates into account. The inverse of standard error squared is also used to weight the trend line shown in Figure 2. To avoid sensitivity to extremely large variance weights, weight values are truncated from above at 100, which corresponds to standard errors on treatment effect estimates of .10 standard deviations.

The effect sizes shown in Figure 2 are estimates and therefore subject to sampling variation. Estimating their variability with a random effects model or with a Bayesian shrinkage estimator reduces the standard deviation across all study-average effects sizes by nearly 50%.

Appendix Table 1. Data Used in Figure 2

| Program name | Starting year | Average effect size at end-of-program | 1 / Study level standard-error squared | Head Start=1 |
|---|---|---|---|---|
| Evaluation of the effect of Head Start program on cognitive growth of disadvantaged children | 1963 | 0.54 | 15.13 | 1 |
| National Head Start and Summer Head Start program, 1965-1968 | 1965 | 0.10 | 53.46 | 1 |
| Summer Head Start in Kearney, NE | 1965 | 0.08 | 6.08 | 1 |
| Summer Head Start in San Jose, CA | 1965 | 0.42 | 14.34 | 1 |
| Summer Head Start in Cambridge, MA | 1965 | 0.14 | 11.03 | 1 |
| Head Start in Miami, FL | 1965 | 0.79 | 9.63 | 1 |
| Summer Head Start in Duluth, MN | 1965 | 0.38 | 14.93 | 1 |
| Impact of 1965 Summer Head Start on children's concept attainment by Allerhand | 1965 | 0.46 | 13.28 | 1 |
| Summer Head Start in Lincoln, NE, matched pairs | 1965 | 0.03 | 71.07 | 1 |
| Summer Head Start in Lincoln, NE, unmatched pairs | 1965 | -0.06 | 3.04 | 1 |
| Bicultural preschool programmed Head Start | 1966 | 0.74 | 6.29 | 1 |
| Summer Head Start in Washington, DC | 1966 | -0.07 | 35.40 | 1 |
| Head Start effects on children's behavior and cognitive functioning one year later by Nummedal and Stern | 1967 | -0.13 | 13.50 | 1 |
| Experiments in Head Start and early education: Direct instruction in Head Start vs. No preschool | 1968 | 0.47 | 12.14 | 1 |
| Experiments in Head Start and early education: Enrichment in Head Start vs. No preschool | 1968 | 0.40 | 13.25 | 1 |
| Head Start in rural areas, MN | 1968 | 0.70 | 15.93 | 1 |
| Comparison of children enrolled in Head start or no preschool in two cities | 1969 | 0.33 | 93.45 | 1 |
| Planned Variation in Head Start, Head Start with curricula vs. no Head Start | 1969 | 0.32 | 10.13 | 1 |
| Planned Variation in Head Start, standard Head Start vs. no Head Start | 1969 | 0.27 | 23.96 | 1 |
| Head Start in New Haven, CT | 1971 | 0.55 | 12.91 | 1 |
| Head Start in New Haven, CT | 1971 | 0.39 | 25.94 | 1 |
| Head Start Bilingual Bicultural Development Project, bilingual Head Start vs. Stay at home | 1979 | 0.29 | 16.18 | 1 |
| Head Start Bilingual Bicultural Development Project, standard Head Start vs. Stay at home | 1979 | 0.33 | 6.99 | 1 |
| Head Start in New Haven, CT | 1979 | 0.55 | 17.33 | 1 |
| Head Start in Guam | 1985 | 0.02 | 19.95 | 1 |
| Early Head Start Research and Evaluation Project | 1996 | 0.13 | 100.00 | 1 |
| ECLS-K Head Start Study, white | 1997 | -0.06 | 100.00 | 1 |
| ECLS-K Head Start Study, African-American | 1997 | 0.05 | 100.00 | 1 |
| ECLS-K Head Start Study, Hispanic | 1997 | -0.04 | 39.42 | 1 |
| Southeastern Head Start program of high quality | 1998 | 0.35 | 24.43 | 1 |
| National Head Start Impact Study First Year, 3-year-old cohort | 2002 | 0.26 | 47.99 | 1 |
| National Head Start Impact Study First Year, 4-year-old cohort | 2002 | 0.25 | 100.00 | 1 |
| NYC Early Head Start, 12 months of younger | 2002 | -0.34 | 1.30 | 1 |

| | | | | |
|---|---|---|---|---|
| NYC Early Head Start, older than 12 months | 2002 | 0.23 | 0.61 | 1 |
| Early Training Project | 1962 | 1.04 | 8.47 | 0 |
| Preschool intervention with disadvantaged children in early 1960s Cambridge, Pre-k & responsive environments | 1962 | -0.01 | 11.42 | 0 |
| Preschool intervention with disadvantaged children in early 1960s Cambridge, Pre-k | 1962 | 0.14 | 9.36 | 0 |
| Preschool intervention with disadvantaged children in early 1960s Cambridge, responsive environments | 1962 | 0.00 | 8.72 | 0 |
| The Perry Preschool | 1962 | 1.10 | 9.23 | 0 |
| Howard University Preschool Program | 1964 | 0.47 | 20.41 | 0 |
| 6-month Full-Day Preschool for 3- to 4-year-olds with curriculum emphasis on language, social skills, & health | 1964 | 0.74 | 10.38 | 0 |
| Karnes Ameliorative Curriculum Pre-K | 1965 | 1.60 | 1.35 | 0 |
| NY Pre-K for disadvantaged children | 1965 | 0.08 | 100.00 | 0 |
| Preschool in Fremont, CA | 1966 | -0.19 | 7.29 | 0 |
| Urbana Infant Home Tutoring Program | 1966 | 1.10 | 6.40 | 0 |
| Greensboro Demonstration Nursery Center | 1967 | 0.81 | 7.29 | 0 |
| Project Know-How | 1967 | 0.65 | 13.16 | 0 |
| Yale Child Welfare Research Program | 1968 | 1.22 | 7.88 | 0 |
| Home Oriented Preschool Education, TV & Home visit & Mobile classroom vs. control | 1968 | 0.15 | 2.88 | 0 |
| Home Oriented Preschool Education, TV & Home visit & Early childhood classroom vs. control, 1969 | 1968 | 0.38 | 6.20 | 0 |
| Home Oriented Preschool Education, TV & Home visit & Early childhood classroom vs. control, 1970 | 1968 | 0.67 | 5.53 | 0 |
| Syracuse University Family Development & Research | 1969 | 0.46 | 100.00 | 0 |
| Abecedarian Project | 1972 | 0.90 | 16.47 | 0 |
| Pittsfield, MA Parent-Child Home Program, 1978 | 1974 | 0.36 | 16.41 | 0 |
| Pittsfield, MA Parent-Child Home Program, 1983 | 1974 | 0.86 | 8.35 | 0 |
| Infants in community-based day care | 1976 | 0.39 | 1.34 | 0 |
| Project CARE | 1978 | 0.44 | 4.27 | 0 |
| Field Teacher Mother Parent Training | 1979 | 0.62 | 17.29 | 0 |
| BYU Pre-school program effects on Low-Risk Children | 1980 | 0.48 | 18.09 | 0 |
| Child care from infancy to 3 years of age | 1981 | 0.01 | 5.71 | 0 |
| Children of Poverty: A multi-level analysis of the determinants of intellectual development | 1983 | 0.74 | 29.98 | 0 |
| Infant Health and Development Program | 1984 | 0.26 | 76.74 | 0 |
| Chicago Parent Center | 1985 | 0.17 | 93.85 | 0 |
| Even Start National Evaluation | 1991 | 0.04 | 27.13 | 0 |
| North Carolina Pre-K Evaluation, Chapter 1 preschool | 1991 | 0.10 | 47.52 | 0 |
| North Carolina Pre-K Evaluation, Family-focused preschool | 1991 | 0.13 | 18.29 | 0 |
| North Carolina Pre-K Evaluation, Community daycare | 1991 | 0.49 | 20.06 | 0 |
| Comprehensive Child Development Program | 1991 | 0.03 | 100.00 | 0 |
| The NICHD Study of Early Child Care and Youth Development | 1991 | 0.20 | 100.00 | 0 |
| Kentucky Educational Reform Act (KERA) Preschool Program, cohort 1 | 1992 | 0.22 | 79.58 | 0 |
| Kentucky Educational Reform Act (KERA) Preschool Program, cohort 3 | 1992 | 0.11 | 63.45 | 0 |

| | | | | |
|---|---|---|---|---|
| Georgia Pre-K Program | 1993 | 0.23 | 53.98 | 0 |
| Michigan School Readiness Program Longitudinal Evaluation | 1995 | 0.25 | 52.17 | 0 |
| Charlotte Bright Beginnings Pre-K Evaluation | 1997 | 0.29 | 100.00 | 0 |
| Center-based language intervention program | 1997 | 1.09 | 2.99 | 0 |
| Parent-child intervention for depressed mothers | 1997 | 0.43 | 20.57 | 0 |
| Literature-Math-Scaffolding Intervention | 1998 | 0.26 | 14.83 | 0 |
| Third Even Start evaluation | 1999 | -0.19 | 38.91 | 0 |
| Utah Pre-K Kindergarten Readiness Program | 2003 | 0.16 | 43.92 | 0 |
| Five State Pre-K RD Evaluation, Michigan | 2004 | 0.13 | 100.00 | 0 |
| Five State Pre-K RD Evaluation, New Jersey | 2004 | 0.09 | 100.00 | 0 |
| Five State Pre-K RD Evaluation, Kolahoma | 2004 | 0.14 | 100.00 | 0 |
| Five State Pre-K RD Evaluation, South Carolina | 2004 | 0.14 | 100.00 | 0 |
| Five State Pre-K RD Evaluation, West Virginia | 2004 | 0.15 | 100.00 | 0 |

**Appendix References** (* indicates that the study contributed data to Figure 2)

*****Abbott-Shim, M., Lambert, R. and 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(2): 191-214.

*Abelson, W.D. (1974). Head Start graduates in school: Studies in New Haven, Connecticut. Ryan, S. (Ed.), *A Report on Longitudinal Evaluations of Preschool Programs*. Volume 1, 1-14. Washington, DC: Office of Child Development, US Department of Health, Education and Welfare.

*Ackerman Ross, F. S. (1985). The relationship of day care to middle-class three-year-olds' language performance (maternal employment, cognitive). Unpublished Ph.D., Memphis State University, United States -- Tennessee.

*Administration for Children and Families (2002). Making a Difference in the Lives of Infants and Toddlers and Their Families: The Impacts of Early Head Start. Washington, DC: U.S. Department of Health and Human Services, 2002.

*Administration on Children, Youth and Families (2001). Building their futures: How Early Head Start programs are enhancing the lives of infants and toddlers in low-income families. Washington, DC: Department of Health and Human Services, 2001.

*Allerhand, M. (1965). Impact of summer 1965 Head Start on children's concept attainment during kindergarten. (ED 015 733)

*Ametjian, A. (1965). The effects of a preschool program upon the intellectual development and social competency of lower class children. Doctoral dissertation, Stanford University, Palo Alto, CA.

*Barnett, W. S., Jung, K., Lamy, C., Wong, V., & Cook, T. (2007). Effects of five-state prekindergarten programs on early learning. Paper presented at the SRCD annual meeting, Boston, MA.

*Beach, D. W. (2004). The effects of a school district's kindergarten readiness summer program on phonological awareness skills of at-risk prekindergarten students: A regression discontinuity analysis. Unpublished Ph.D., Utah State University, United States -- Utah.

*Bennett, T. L. (2000). Teachers' use of children's literature, mathematics manipulatives, and scaffolding to improve preschool mathematics achievement: Does it work? United States -- Texas, University of North Texas.

*Bertram, C. L., Hines, B.W., & MacDonald, R. (1971). Summative Evaluation of the Home-Oriented Preschool Education Program: Summary Report. Division of Research and Evaluation, Appalachia Educational Laboratory.

*Bissell, J. S. (1973). Bissell Planned variation in Head Start and Follow Through J.S. Stanley (Ed.), *Compensatory education for children ages two to eight*, Johns Hopkins University Press, Baltimore (1973), pp. 63–107

*Blatt, B. & Garfunkel, F. (1969). *The educability of intelligence: Preschool intervention with disadvantaged children.* Washington, DC:  The Council for Exceptional Children Inc.

Borenstein M., Hedges, L., Higgins, J., & Rothstein, H. (2005). Comprehensive Meta-analysis, Version 2. Englewood N. J.: Biostat.

*Bryant, D. M., Peisnar-Feinberg, E. S., and Clifford, R. M. (1993). Evaluation of Public Preschool Programs in North Carolina, University of North Carolina (Chapel Hill, N.C.: 1993)

*Bridge, C. A., Townley, K. F., Hemmeter, M. L. & de Mesquita, P. B. (1993). Third party evaluation. Kentucky Education Reform Act (KERA) preschool programs. Lexington: University of Kentucky, College of Education and College of Human Environmental Sciences.

*Bridge, C. A., Townley, K. F., Hemmeter, M. L. & de Mesquita, P. B. (1995). Third party evaluation of the Kentucky Education Reform Act preschool programs. Frankfort, KY: Kentucky Department of Education. (ED 394 628)

*Brooks-Gunn, J., Liaw, F., & Klebanov, P. (1992). Effects of early intervention on low birth weight preterm infants: What aspects of cognitive functioning are enhanced? *Journal of Pediatrics*, 120(3), 350-359.

*Brooks-Gunn, J., McCarton, C. M., Casey, P. H., et al. (1994). Early intervention in low-birth-weight premature infants: Results through age 5 years from the infant health and development program. *Journal of the American Medical Association*, 272(16), 1257-1262.

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).

*Chesterfield, R. et al (1982). An Evaluation of the Head Start Bilingual Bicultural Curriculum Development Project. Final Report. (ERIC Document Reproduction Service No. ED 212 391)

*Cicarelli, V. G., Cooper, W. H. & Granger, R. L. (1969). The impact of Head Start: An evaluation of the effects of Head Start on children's cognitive and affective development.

Volume 2. Office of Economic Opportunity. Athens, OH: Westinghouse Learning Corporation and Ohio University. PB 184 329.

*Cline, M. G. & Dickey, M. (1968). An evaluation and follow-up study of summer 1966 Head Start children in Washington, DC. Washington, DC: Howard University. (ED 020 794)

*Di Lorenzo, L. T., Salter, R. & Brady, J. J. (1969). Prekindergarten programs for educationally disadvantaged children. Final report. New York: New York State Department of Education. (ED 038 460)

*Dunham, R., Kidwell, J., & Portes, P. (1988). Effects of parent-adolescent interaction on the continuity of cognitive development from early childhood to early adolescence. *Journal of Early Adolescence*, 8(3), 297-310.

*Erickson, E. L., McMillan, J., Bonnell, J., Hofman, L. & Callahan, O.D. (1969). Experiments in Head Start and early education: The effects of teacher attitude and curriculum structure on preschool disadvantaged children. Final report. 186. Office of Economic Opportunity, Washington, DC. Washington, DC. Western Michigan University, Kalamazoo. Center for Sociological Research. OEO-4130. (ED 041 615)

*Esteban, M. (1987). A comparison of Head Start and non-Head Start reading readiness scores of low-income kindergarten children of Guam. Dissertation Abstracts International, 49, 0723. (UMI No. 8808677)

*Field, T., Pickens, J., et al. (2000). "Targeting adolescent mothers with depressive symptoms for early intervention." *Adolescence* 35(138): 381-414.

*Field, T., Widmayer, S., Greenberg, R., & Stoller, S. (1982). Effects of parent training on teenage mothers and their infants. *Pediatrics*, 6, 703–707.

*Gray, S. W., & Klaus, R. A. (1970). The early training project: A seventh-year report. *Child Development*, 41, 909–924.

*Henderson, R. W., Rankin, R. J. & Frosbisher, M. W. (1969). Positive effects of a bicultural preschool program on the intellectual performance of Mexican-American children. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles: Tucson: Arizona University, College of Education.

*Herzog, E., Newcomb, C. H. & Cisin, I. H. (1973). Preschool and postscript: An evaluation of the inner-city program. Washington, DC: Social Research Group, Washington University. (ERIC Document Reproduction Service ED 118241)

*Hines, B. W. (1971). Analysis of Visual Perception of Children in the Appalachia Preschool Education Program. Division of Research and Evaluation, Appalachia Educational Laboratory.

*Hines, B. W. (1971). Attainment of Cognitive Objectives. Division of Research and Evaluation, Appalachia Educational Laboratory.

*Hines, B. W. (1971). Detailed Analysis of Language Development of Preschool Children in ECE Program. Technical Report No. 4. Division of Research and Evaluation, Appalachia Educational Laboratory.

*Howard, J. L. and Plant, W. T. (1967). "Psychometric evaluation of an operant Head Start program." *Journal of Genetic Psychology* 111: 281-288.

*IHDP research team (1990). Enhancing the outcomes of low-birth-weight, premature infants: A multisite, randomized trial. *The Journal of the American Medical Association*, 263, 3035-3042.

Jacob, R. T., Creps, C. L., & Boulay, B. (2004). *Meta-analysis of research and evaluation studies in early childhood education.* Cambridge, MA: Abt Associates Inc.

*Karnes, M. B., Hodgins, A. S., & Teska, J. A. (1969). The effects of early education with disadvantaged infants. In Investigations of classroom and at-home interventions: Research and development programs on preschool disadvantaged children.

*Karnes, M. B., Hodgins, A. S., Stoneburner, R. L., Studley, W. M. & Teska, J. A. (1968). "Effects of a highly structured program of language development on intellectual functioning and psycholinguistic development of culturally disadvantaged three-year-olds." *The Journal of Special Education* 2(4): 405-412.

*Keister, M., (1978). A demonstration project: Group care for infants and toddlers. Final Report, University of North Carolina at Greensboro.

*Klebanov, P. K., Brooks-Gunn, J., & McCormick, M. C. (2001). Maternal coping strategies and emotional distress: results of an early intervention program for low birth weight young children. *Developmental Psychology*, 37(5), 654-667.

*Krider, M. A. & Petsche, M. (1967). An evaluation of Head Start pre-school enrichment programs as they affect the intellectual ability, the social adjustment, and the achievement level of five-year-old children enrolled in Lincoln, Nebraska. Lincoln, NE: University of Nebraska. (ERIC Document Reproduction Service No. ED 015 011)

*Lally, J. R. & Honig, A. S. (1977). The Family Development Research Program: A program for prenatal, infant & early childhood enrichment. Final Report. Syracuse, NY: Syracuse University. (ED 143 458)

*Larsen, J. M. (1985). Family influences on competence in low-risk preschool children. Toronto, Canada: Society for Research in Child Development.

*Larson, D. F. (1969). The effects of a preschool experience upon intellectual functioning among four-year-old, white children in rural Minnesota. Mankato: Minnesota State University, College of Education. (ED 039 030)

Layzer, J. I., Goodson, B. D., Bernstein, L., & Price, C. (2001). *National evaluation of family support programs, volume A: The meta-analysis, final report.* Cambridge, MA: Abt Associates Inc.

*Lee, M., & Burchinal, M. (1987). Children of Poverty: A Multi-Level Analysis of the Determinants of Intellectual Development.

*Lee, V. E., Schnur, E. & Brooks-Gunn, J. (1988). Does Head Start work? A 1-year follow-up comparison of disadvantaged children attending Head Start, no preschool, and other preschool programs. *Developmental Psychology* 24(2): 210-222.

*MacDonald, R. (1971). Analysis of Intelligence Scores. Division of Research and Evaluation, Appalachia Educational Laboratory.

*McNamara, J.R. (1968). Evaluation of the effects of Head Start experience in the area of self-concept, social skills, and language skills. Pre-publication draft. Miami, FL: Dade County Board of Public Instruction. (ED 028 832)

*Morris, B. & Morris, G.L. (1966). Evaluation of changes occuring in children who participated in project Head Start. Kearney, NE: Kearney State College. (ED 017 316)

*NICHD Early Child Care Research Network & Duncan, G. J. (2003). Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development*, 74, 1454–1475.

*Nummedal, S. G. & Stern, C. (1971). Head Start graduates: One year later. Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

*Pilcher, L. C. (1994). Georgia Prekindergarten Program Evaluation with Executive Summary. Atlanta, GA: Georgia State University Early Childhood Education.

*Porter, P. J., Leodas, C., Godley, R. A. & Budroff, M. (1965). Evaluation of Head Start educational program in Cambridge, Massachusetts. Final Report. Cambridge, MA: Harvard University. (ED 013 668)

*Ramey, C. T., & Campbell, F. A. (1984). Preventive education for high-risk children: Cognitive consequences of the Carolina Abecedarian Project. *American Journal of Mental Deficiency*, 88, 515–523.

*Rescorla, L. A., Provence, S., & Naylor, A. (1982). The Yale Child Welfare Research Program: Description and results. In E. F.Zigler & E.Gordon (Eds.), *Day care: Scientific and social policy issues* (pp. 183–199). Boston: Auburn House.

*Reynolds, A. J. (1995). One year of preschool intervention or two: Does it matter? *Early Childhood Research Quarterly*, 10, 1–31.

*Roberts, J. E., Rabinowitch, S., Bryant, D. M., Burchinal, M. R., Koch, M. A. & Ramey, C. T. (1989). Language skills of children with different preschool experiences. *Journal of Speech, Language, and Hearing Research* 32(4): 773-786.

*Robertson, S. B. & S. E. Weismer (1999). Effects of treatment on linguistic and social skills in toddlers with delayed language development. *Journal of Speech, Language, and Hearing Research* 42(5): 1234-1248.

*Rubenstein, J. L.. (1981). A Two-Year Follow-Up of Infants in Community-Based Day Care. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 22(3), 209.

*Smith, E. J., Pellin, B. J., Agruso, S. A., & Educational Research Service, A. V. A. (2003). Bright Beginnings: An Effective Literacy-Focused PreK Program for Educationally Disadvantaged Four-Year-Old Children (No. 1-931762-15-5).

*Sontag, M., Sella, A. P. & Thorndike, R. L. (1969). The effect of Head Start training on the cognitive growth of disadvantaged children. *The Journal of Educational Research* 62(9): 387-389.

*St. Pierre, R., Swartz, J., Gamse, B., Murray, S., Deck., D. & Nickel, P. (1993) National Evaluation of the Even Start Family Literacy Program, Final report. Washington, DC: US Department of Education.

*St. Pierre, R. G., Layzer, J. I., Goodson, B. D., Bernstein, L. S. (1997). National Impact Evaluation of the Comprehensive Child Development Program. Cambridge, MA: Abt Associates, Inc.

*St. Pierre, R. G., Ricciuti, A. E., & Rimdzius, T. A. (2005). Effects of a Family Literacy Program on Low-Literate Children and Their Parents: Findings From an Evaluation of the Even Start Family Literacy Program. Developmental Psychology, 41(6), 953-970.

*Ricciuti, A. E, St. Pierre., R. P., Lee, W., Parsad, A., Institute of Education Sciences, W. D. C., & Abt Associates, (2004). Third National Even Start Evaluation: Follow-Up Findings from the Experimental Design Study. NCEE 2005-3002: National Center for Education Evaluation and Regional Assistance NCEE.

*Tamminen, A. W., Weatherman, R. F. & McKain, C. W. (1967). An evaluation of a preschool training program for culturally deprived children. Final report. U.S. Department of Health, Education, and Welfare. Duluth: University of Minnesota. BR-5-8156. (ED 019 135)

*U.S. Department of Health and Human Services, Administration for Children and Families (2005), Head Start Impact Study: First Year Findings, Washington, DC.

*Vance, B. J. (1967). The effect of preschool group experience on various language and social skills in disadvantaged children. Final report. (ED 019 989)

*Vele, E. M. (2004). Relationship-based, dyadic Early Head Start intervention. Unpublished Ph.D., Yeshiva University, United States -- New York.

*Weikart, D. P., Bond, J. T. & McNeil, J. T. (1978). The Ypsilanti Perry Pre-School Project: Pre-school years and longitudinal results through fourth grade. Monographs of the High/Scope Educational Research Foundation 3. Ypsilanti, MI: High/Scope Educational Research Foundation.

*Xiang, Z. & Schweinhart, L. J. (2002). Effects five years later: The Michigan School Readiness Program evaluation through age 10. Michigan State Board of Education. Ypsilanti, MI: High/Scope Educational Research Foundation.

*Young-Joo, K. (2007). The return to private school and education-related policy. University of Wisconsin (dissertation).

*Zigler, E. F., Ableson, Trickett & Seitz (1982). Is an intervention program necessary in order to improve economically disadvantaged children's IQ scores? *Child Development* 53(2): 340-348.

ENDNOTES

[1] The federal government also provides some financial assistance to families seeking child care via the Child and Dependent Care Tax Credit and exclusions from income for benefits under dependent care assistance programs; however, few low-income families benefit from these programs (Greenburg, 2007; Magnuson et al., 2007).

[2] As described in the appendix, programs selected for our analysis had both treatment and control/comparison groups, included at least 10 participants in each condition, incurred less than 50% attrition and measured children's cognitive development close to end of their "treatment" programs. Studies had to have used random assignment or one of the following quasi-experimental designs: change models, fixed effects modes, regression discontinuity, difference in difference, propensity score matching, interrupted time series, instrumental variables and some other types of matching. Studies that used quasi-experimental designs must have had pre- and post-test information on the outcome or established baseline equivalence of groups on demographic characteristics determined by a joint-test.

[3] A third explanation would be that program impacts do not persist because early elementary instruction is most beneficial to children who enter school with high levels of initial skills, and that Head Start program impacts are not sufficiently large enough to get children to a point at which they will benefit from such instruction. There does not seem to be good evidence to support this conjecture.

[4] This estimate comes from Duncan and Sojourner (forthcoming) and is based on weights designed to match the demographic characteristics of the IHDP sample to those of all U.S. births.

[5] None of these studies is based on random assignment of children to different pre-school class sizes, nor do any conduct long-run follow-ups. Chetty et al. (2011) find noteworthy longer-run impacts of assignment to smaller Kindergarten to grade 3 classrooms in the Project Star data.