



Published in final edited form as:

*Dev Psychol.* 2014 December ; 50(12): 2572–2586. doi:10.1037/a0038205.

## Head Start's Impact is Contingent on Alternative Type of Care in Comparison Group

**Fuhua Zhai, Ph.D. [Associate Professor],**

Fordham University Graduate School of Social Service, 113 West 60th Street, New York, NY 10023, Phone: 646-293-3966

**Jeanne Brooks-Gunn, Ph.D. [Professor], and**

Columbia University Teachers College and the College of Physicians and Surgeons, 525 West 120th Street, New York, NY 10027, Phone: 212-678-3369

**Jane Waldfogel, Ph.D. [Professor]**

Columbia University School of Social Work, 1255 Amsterdam Avenue, New York, NY 10027, Phone: 212-851-2408

Fuhua Zhai: fzhai1@fordham.edu; Jeanne Brooks-Gunn: brooks-gunn@columbia.edu; Jane Waldfogel: jw205@columbia.edu

### Abstract

Using data ( $n = 3,790$  with 2,119 in the 3-year-old cohort and 1,671 in the 4-year-old cohort) from 353 Head Start centers in the Head Start Impact Study, the only large-scale randomized experiment in Head Start history, this paper examined the impact of Head Start on children's cognitive and parent-reported social-behavioral outcomes through first grade contingent on the child care arrangements used by children who were randomly assigned to the control group (i.e., parental care, relative/non-relative care, another Head Start program, or other center-based care). A principal score matching approach was adopted to identify children assigned to Head Start who were similar to children in the control group with a specific care arrangement. Overall, the results showed that the effects of Head Start varied substantially contingent on the alternative child care arrangements. Compared to children in parental care and relative/non-relative care, Head Start participants generally had better cognitive and parent-reported behavioral development, with some benefits of Head Start persisting through first grade; in contrast, few differences were found between Head Start and other center-based care. The results have implications regarding the children for whom Head Start is most beneficial as well as how well Head Start compares to other center-based programs.

### Keywords

Head Start; center-based care; parental care; principal score matching; Head Start Impact Study

---

Head Start has been the single largest publicly financed early childhood education and care program in the U.S. since its creation in 1965 as part of the War on Poverty. It aims to

improve the school readiness of low-income preschool-age children, particularly 3- and 4-year olds, by providing high-quality and comprehensive early education and other services. Whether Head Start has been effective has been debated, in large part because until recently no randomized trial of Head Start had been conducted (Besharov & Call, 2009; Nisbett, 2009; Styfco & Zigler, 2004; see Camilli, Vargas, Ryan, & Barnett, 2010, for a meta-analysis of randomized evaluations of other preschool education programs). One challenge for the non-experimental Head Start studies, even those using the most rigorous methods (e.g., Currie & Thomas, 1995, 1999; Deming, 2009; Garces, Thomas, & Currie, 2002; R. Lee, Zhai, Brooks-Gunn, Han, & Waldfogel, 2014; Ludwig & Miller, 2007; Zhai, Brooks-Gunn, & Waldfogel, 2011), is to account adequately for selection bias, given that the program by design serves children who are economically disadvantaged and tend to have worse developmental outcomes than their more advantaged counterparts even before attending Head Start.

In the 1998 reauthorization of Head Start, Congress mandated that the U.S. Department of Health and Human Services (USDHHS) determine the impact of Head Start on the children it serves. Under this legislative mandate, the Head Start Impact Study (HSIS) selected a nationally representative sample of 3- and 4-year-old Head Start-eligible children and randomly assigned them to Head Start or control conditions. As the only large-scale randomized experiment in Head Start history, the HSIS reported short-term benefits of Head Start in multiple domains one year after random assignment, although few of these benefits were sustained through kindergarten or first or third grade (USDHHS, 2005, 2010, 2012).

However, because children in the HSIS experiment were randomized only to a treatment or control condition, the HSIS experimental analysis could not address the question of whether the effects of Head Start varied depending on the types of child care arrangements that children in the control group selected. To address this question, non-experimental methods must be used.

In this analysis we used data from the HSIS to investigate how the effects of Head Start on children's cognitive and social-behavioral outcomes varied contingent on the counterfactual child care arrangements to which it was compared by year and by cohort. Using a principal score approach (similar to that used by Hill, Waldfogel, & Brooks-Gunn, 2002), we compared Head Start in the treatment group to specific care arrangements that children received in the control group, including parental care, relative/non-relative care, other Head Start programs, and other center-based care. By taking into account variation in the counterfactual, this analysis provided a more detailed picture of how well Head Start worked for Head Start eligible children compared to other specific child care arrangements. The findings may have important implications for policymakers, in particular with regard to targeting children who otherwise would attend the least beneficial arrangements if they were not enrolled in Head Start and with regard to how well Head Start compares to other center-based programs.

## Background and Literature Review

A common challenge in Head Start research has been how to account adequately for the issue of selection. Disadvantaged children are more likely than their more advantaged peers to attend Head Start and also to have worse developmental outcomes (Currie, 2005; Reid, Webster-Stratton, & Baydar, 2004). To address this issue of selection, a number of observational studies have used rigorous statistical methods (e.g., family fixed effects, regression discontinuity, and propensity score matching) and found modest and significant short- and long-term benefits of Head Start (e.g., Currie & Thomas, 1995, 1999; Deming, 2009; Garces et al., 2002; R. Lee et al., 2014; Ludwig & Miller, 2007; Zhai et al., 2011).

To more conclusively address the issue of selection bias, the HSIS randomly assigned a nationally representative sample of 3- and 4-year-old children whose families applied to over-subscribed Head Start programs to either have access to Head Start (i.e., the treatment group) or to be placed on a waiting list (i.e., the control group). One year after random assignment, the HSIS reported significant benefits of Head Start in multiple domains (USDHHS, 2005). For example, compared to children in the control group of the same age cohort, children in the treatment group had significantly better cognitive development as measured by higher scores on the Peabody Picture Vocabulary Test (PPVT) Receptive Vocabulary and the Woodcock-Johnson III (WJ-III) subscales in Letter-Word Identification and math in Applied Problems (effect sizes, calculated by regression coefficients divided by standard deviations of the measures, were 0.18, 0.26, and 0.15 in the 3-year-old cohort; and 0.09, 0.22, and not significant in the 4-year-old cohort, respectively). Children of the 3-year-old cohort in the treatment group also had lower scores on Hyperactive Behavior (effect sizes of 0.21) than their peers in the control group (not significant in the 4-year-old cohort). Nevertheless, few of these benefits persisted through kindergarten or thereafter (USDHHS, 2010, 2012).

These estimated effects of Head Start on children's development, especially cognitive outcomes after one-year participation in the HSIS, were smaller compared to those reported in earlier evaluations of model early interventions (e.g., Perry Preschool, Abecedarian, and the Infant Health and Development Program [IHDP]; with short-term effect sizes of 0.35–0.97 on cognitive outcomes) (Brooks-Gunn, 2011; Camilli et al., 2010; Karoly, Kilburn, & Cannon, 2005; Ludwig & Phillips, 2007). This difference may reflect the fact that the counterfactual has changed, given that few 3- and 4-year-old children in the 1960s to the 1980s attended any form of preschool if they did not have Head Start or other model early interventions evaluated in many prior studies while most of these age-groups of children today have some form of school- or center-based care (Waldfogel, 2006). The non-compliance rates in the HSIS (i.e., 15% and 22% for no-shows of children who were assigned to the treatment group but did not attend Head Start and 16% and 13% for crossovers of children who were assigned to the control group but attended Head Start in the 3- and 4-year-old cohorts, respectively) were relatively high compared to those in the prior model early interventions, which may also indicate the increase in the number of preschool programs across the U.S. in the past decades.

However, in many prior studies the counterfactual of child care arrangements to which Head Start is compared has not been clearly defined or directly examined. Children who do not attend Head Start are in a variety of alternative care settings, ranging from exclusive parental care to informal relative or non-relative child care to other high-quality early education programs (R. Lee et al., 2014; V. Lee, Brooks-Gunn, & Schnur, 1988; USDHHS, 2005, 2010; Zhai et al., 2011). For example, in our analysis sample of the HSIS, in the control group 41% of the 3-year-old cohort and 38% of the 4-year-old cohort received only parental care, 18% and 11% respectively received relative/non-relative care in the child's home or another home, 25% and 37% respectively received other center-based care, and 16% and 13% respectively attended another Head Start program (i.e., a Head Start program that was not part of the experimental study). Prior research has shown that the type and quality of child care arrangements are associated with children's developmental outcomes (Baydar & Brooks-Gunn, 1991; Gormley, 2008; Magnuson, Ruhm, & Waldfogel, 2007; NICHD Early Child Care Research Network [ECCRN], 2005; NICHD ECCRN & Duncan, 2003; Smolensky & Gootman, 2003; Waldfogel, 2006).

Therefore, a comparison of Head Start to the specific care arrangements received by children in the control group would be informative, as shown by Hill et al. (2002) in the analysis of the IHDP, a randomized controlled trial (RCT) of early childhood intervention services provided to low birth weight premature children in eight sites across the nation. Using a principal score matching method, Hill and colleagues found that the IHDP program had the largest benefits for children who otherwise would have received parental care. Similar findings on Head Start were obtained in two recent observational studies (R. Lee et al., 2014; Zhai et al., 2011), which used data from the Early Childhood Longitudinal Study–Birth Cohort (ECLS–B) and the Fragile Families and Child Wellbeing Study (FFCWS), respectively, and found that Head Start was associated with improved cognitive outcomes compared to parental care or relative/non-relative care, but was not different from other center-based care.

## The Present Study

This study used data from the HSIS to examine the effects of Head Start on children's cognitive and parent-reported social-behavioral outcomes by comparing Head Start in the treatment group to specific child care arrangements for children in the control group, including parental care, relative/non-relative care, other Head Start programs, and other center-base care. We analyzed the effects of Head Start by cohort and by year from Head Start year through first grade. Following Hill et al. (2002), a principal score approach was used to identify subgroups of children in the treatment group who, without the intervention, would have had child care arrangements similar to subgroups of children in the control group.

Based on the findings from the randomized trial of IHDP using a similar procedure (Hill et al., 2002) and the non-experimental analyses of two nationwide data sets using propensity score matching (R. Lee et al., 2014; Zhai et al., 2011), we would expect Head Start to have the largest cognitive benefits when compared to parental care or relative/non-relative care. The findings on the associations between child care arrangements and social-behavioral

outcomes have been mixed in the literature. For example, some observational studies found that children who attended center-based care tended to have more behavior problems in preschool and elementary school years than children who had parental or relative care (Magnuson & Waldfogel, 2005; NICHD ECCRN, 2005). In contrast, studies using RCT designs found reduced, or no elevated, behavior problems among children who attended high-quality early interventions (Hill et al., 2002; Love, Chazan-Cohen, Raikes, & Brooks-Gunn, 2013). Since randomized assignment was used in the HSIS and Head Start generally provides comprehensive services to participants, we would expect Head Start to be associated with increased social skills and reduced behavior problems when compared to parental care or relative/non-relative care. In contrast, given the variation in the quality of both Head Start and other center-based care programs (e.g., USDHHS, 2005), we would expect to find few differential effects on cognitive or social-behavioral outcomes in the comparisons of Head Start in the treatment group to Head Start or other center-based care in the control group. Since both the 3- and 4-year-old cohorts in the treatment group received Head Start, we would expect similar initial effects of Head Start by the end of the first Head Start year. Prior research found evidence that the length of participation (or program dosage) in high-quality interventions was associated with better developmental outcomes of children (Hill, Brooks-Gunn, & Waldfogel, 2003; Zhai et al., 2010). Therefore, we would expect to find sustained effects in the 3-year-old cohort, in which most children in the treatment group attended Head Start for two years (USDHHS, 2010), more likely than in the 4-year-old cohort.

## Method

### Data and Analysis Sample

We used the restricted-use data from the HSIS for the analyses. Under the congressional mandate in 1998, the HSIS included a nationally representative sample of newly entering, Head Start eligible 3- and 4-year-old children who were randomly assigned to either the treatment group that had access to Head Start or to the control group that could enroll in other early childhood programs or child care, including parental care. Data were collected from preschool to third grade between fall 2002 and spring 2008 (see USDHHS, 2005, 2010, 2012, for the detailed procedures of design, sampling, random assignment, and data collection in the HSIS). Data through first grade were used in this analysis.

The original HSIS sample included 4,667 children from 383 randomly selected Head Start centers in 84 randomly selected grantees/delegate agencies spread over 23 different states. Children from Puerto Rico ( $n=225$ ) were excluded from the restricted-use HSIS data. Among the 4,442 children in the HSIS restricted-use data (2,646 in the treatment group and 1,796 in the control group), 3,790 children (2,357 in the treatment group and 1,433 in the control group) from 353 Head Start centers had non-missing information on focal child care arrangements in spring 2003 (i.e., one year after random assignment)<sup>1</sup>.

---

<sup>1</sup>The results of *t*-tests by random assignment status and cohort found few significant differences in demographics and family background between children included in and those excluded from our analyses. Limited evidence suggests that children included in our analyses had higher levels of household risk than children who were excluded.

Consistent with the main analyses in the HSIS reports, we analyzed children in the 3-year-old cohort ( $n = 2,119$ ) and the 4-year-old cohort ( $n = 1,671$ ) separately. As shown in Table 1, the analysis data include those collected in fall 2002 (baseline), spring 2003 (one year after random assignment), spring 2004 (second Head Start year for the 3-year-old cohort and kindergarten for the 4-year-old cohort), spring 2005 (kindergarten for the 3-year-old cohort and first grade for the 4-year-old cohort), and spring 2006 (first grade for the 3-year-old cohort and no data collected in the 4-year-old cohort). Table 1 also shows the sample size in analysis among children with non-missing data of child care arrangements in spring 2003 as well as outcome measures by cohort and by year of data collection. For example, in the analysis of outcome variables in spring 2003, the full sample size was 3,778, including 2,111 children in the 3-year-old cohort and 1,667 children in the 4-year-old cohort. In all analyses, we used sampling weights provided in the HSIS data (incorporated with jackknife replicate weights for variance estimation), which were adjusted for non-response in data collection to represent the national population of newly entering Head Start participants for 2002 (USDHHS, 2005).

### Outcome Measures

The outcome measures in this study included children's cognitive and social-behavioral outcomes from the first Head Start year through first grade. Information on children's cognitive development was collected from direct assessment, including the PPVT Receptive Vocabulary and the WJ-III subscales of Letter-Word Identification and math in Applied Problems. Information on social and behavioral development was reported by parents, including Social Skills and Positive Approaches to Learning, Aggressive Behavior, and Hyperactive Behavior. The description and psychometric information on these outcome measures below are based on the HSIS reports (USDHHS, 2005, 2010, 2012).

Among the directly assessed cognitive measures, the HSIS used a shortened version of PPVT ( $\alpha=0.62$  in 3-year-old cohort and  $\alpha=0.79$  in 4-year-old cohort in spring 2003) that was developed to reduce the testing burden imposed on young children, using maximum likelihood Item Response Theory (IRT). The PPVT measures the child's receptive vocabulary, which is listening comprehension for the spoken word in standard English (Dunn, Dunn, & Dunn, 1997). In the assessment, the child points to the one out of four pictures that best represents the meaning of the stimulus word presented orally by the assessor. The WJ-III (Woodcock, McGrew, & Mather, 2001) Letter-Word Identification ( $\alpha=0.87$  in 3-year-old cohort and  $\alpha=0.90$  in 4-year-old cohort in spring 2003) measures letter and word identification skills, including symbolic learning or the ability to match a rebus with an actual picture of the object and reading identification skills in identifying isolated letters and words as they appear in the test easel. The WJ-III Applied Problems scale ( $\alpha=0.89$  in 3-year-old cohort and  $\alpha=0.90$  in 4-year-old cohort in spring 2003) measures the child's ability to analyze and solve practical math problems, including counting and simple calculations.

The HSIS also included parent-reported measures of children's social and behavioral development based on a modified Classroom Behavior Checklist (CBCL; Achenbach, Edelbrock, & Howell, 1987). The scale of Social Skills and Positive Approaches to Learning

includes seven items ( $\alpha=0.61$  in 3-year-old cohort and  $\alpha=0.64$  in 4-year-old cohort in spring 2003), focusing on cooperative and empathic behavior as well as positive approaches to learning such as “Makes friends easily,” “Comforts or helps others,” “Likes to try new things,” and “Shows imagination in work and play.” One of the problem behavior subscales<sup>2</sup> is Aggressive Behavior, which has four items ( $\alpha=0.61$  in 3-year-old cohort and  $\alpha=0.56$  in 4-year-old cohort in spring 2003) of aggressive or defiant behavior such as “Hits and fights with others,” “Has temper tantrums or hot temper,” and “Is disobedient at home.” The Hyperactive Behavior subscale includes three items ( $\alpha=0.62$  in 3-year-old cohort and  $\alpha=0.59$  in 4-year-old cohort in spring 2003) of inattentive or hyperactive behavior such as “Can’t concentrate, can’t pay attention for long” and “Is very restless and fidgets a lot.”

### Measures of Child Care Arrangements and Baseline Covariates

We used the focal child care arrangements as defined in the HSIS reports (USDHHS, 2005, 2010, 2012). For children in either the treatment or control group who were enrolled in Head Start, Head Start was always defined as the focal setting. For children in other arrangements (including those in multiple arrangements) that lasted at least 5 hours per week, the priority of coding the focal settings followed a hierarchical order of other center-based program, non-relative’s home, relative’s home, non-parental care in the child’s own home by a non-relative, and non-parental care in the child’s own home by a relative. The focal setting was parental care if children did not receive non-parental care for more than 5 hours per week. Relatively few children received non-parental, non-center-based care, including relative or non-relative care in the child’s home or another home. Therefore, we combined these focal settings into a category of relative/non-relative care. As a result, the child care arrangements in our analyses had four mutually exclusive categories, including Head Start, other center-based care, relative/non-relative care, and parental care.

In all the regression analyses, as detailed below, we controlled for the same set of baseline covariates<sup>3</sup> that were used in the main or subgroup analyses in the HSIS reports. These child and family covariates, collected at baseline in fall 2002, could not have been influenced by Head Start participation but may have affected subsequent child outcomes. Prior research has demonstrated the importance of accounting for the potential confounding variables of child and family demographic characteristics in detecting the effects of preventive interventions, especially those targeting low-income children (see for example, Aber, Brown, & Jones, 2003; Hill et al., 2002, 2003; R. Lee et al., 2014; Love et al., 2013; Raver et al., 2009; Tolan, Gorman-Smith, & Henry, 2004; Zhai et al., 2011). These covariates may also have affected the compliance with HSIS random assignment for families in the treatment group as well as the selection of child care arrangements for families in the control group. Therefore, including these covariates may increase the precision of analyses and their

<sup>2</sup>The HSIS data also included a Withdrawn Behavior subscale containing three items of shy, withdrawn or depressed behavior. However, the reliability of this measure was quite low ( $\alpha=0.41$  in 3-year-old cohort and  $\alpha=0.38$  in 4-year-old cohort in spring 2003).

<sup>3</sup>Sensitivity tests including additional covariates such as parent-reported depressive symptoms showed similar results to those reported below. Many of these additional covariates had missing data (e.g., 19% on parent-reported depressive symptoms) and were correlated with those included in the models. Using the same set of covariates as in the HSIS reports may help avoid statistical issues such as multicollinearity and lack of common support in the matching process, as detailed below, and also keep consistent with the analyses in the HSIS reports. In addition, we did not control for variables collected after fall 2002, because they could either be subsequent treatments (e.g., second-year Head Start attendance in the 3-year-old cohort) or have been affected by child care arrangements.

explanatory power to detect any true Head Start impacts on the outcomes of interest (USDHHS, 2010).

Specifically, child covariates included gender, race/ethnicity (White/Other, Black, or Hispanic), test language (English vs. Spanish/other) at baseline, and whether the child had special needs at baseline. In addition, following the HSIS reports, the models also included controls for age at spring assessment (in weeks). The HSIS also had measures of children's cognitive and behavioral outcomes administered in the fall 2002. In principle, we would want to include these measures in the models as controls for children's pretreatment scores. However, it should be noted that since most of the fall 2002 data were collected in a 3-month period (i.e., October to December 2002) after random assignment (i.e., May to September 2002), children's initial cognitive measures might have been affected by Head Start participation. Preliminary regression analyses that included covariates as well as sampling weights and jackknife replicate weights showed that Head Start did have some significant effects on the outcomes collected in the fall 2002, especially when compared to the specific care arrangements in the 4-year-old cohort. Therefore, we did not control for these measures in the analyses, assuming children had similar pretreatment scores as a result of random assignment.

As in the analyses in the HSIS reports, we also controlled for parent and family covariates, including mother's age as of 9/1/2002, whether both biological parents lived with child, whether biological mother was a recent immigrant, primary language spoken at home (English vs. Spanish/Other), household risk (low/no, medium, or high, as indexed by five risk factors including receipt of TANF or Food Stamps, neither parent having high school diploma or a GED, neither parent being employed or in school, the child's biological mother being a single parent, and mother giving birth to the child as a teenager), and urbanicity.

### Analytic Strategy

The analyses were conducted by cohort in respective years from Head Start year (spring 2003) through first grade (spring 2006). To replicate the estimates in the HSIS reports, intent-to-treat (ITT) analyses were first conducted for overall comparisons between children who were randomly assigned to the HSIS treatment and control groups (USDHHS, 2005). Treatment-on-treated (TOT) estimates for those who participated in Head Start were then calculated using the same approach employed by the HSIS (USDHHS, 2010), achieved by dividing ITT estimates by  $(1 - n - c)$ , where  $n$  is the rate of no-shows and  $c$  is the rate of crossovers. Following the procedures and strategies adopted by the analyses in the HSIS reports (USDHHS, 2005, 2010, 2012), ordinary least squares (OLS) regressions were conducted for ITT estimates and then TOT estimates were calculated separately in 3- and 4-year-old cohorts by year of data collection. All models incorporated sampling weights and jackknife replicate weights provided in the HSIS data for each wave of the outcome measures, which were adjusted for non-response to represent the national population of newly entering Head Start participants for 2002 (USDHHS, 2010, 2012).

To examine the effects of Head Start compared to the specific child care arrangements of children in the control group, OLS regressions were first conducted in sub-samples containing children in the treatment group who participated in Head Start and children with



a specific care arrangement in the control group. A principal score matching approach was then adopted to identify a group of Head Start participants in the HSIS treatment group who were similar to children with a specific care arrangement in the control group and who would have been most likely to have chosen this care arrangement if they had been assigned to the control group. As a derivative of propensity score matching, principal score matching builds on methodological innovations in principal stratification in the context of randomized experiments (Barnard, Frangakis, Hill, & Rubin, 2003; Frangakis & Rubin, 2002; Hill et al., 2002; Zhai et al., 2010).

Specifically, the principal score matching method was conducted in three stages in the 3- and 4-year-old cohorts separately. In the first stage, child and family covariates, as detailed above, were used to predict the probability of choosing different care arrangements (i.e., Head Start, other center-based care, relative/non-relative care, and parental care) for each child in the control group, using a multinomial logistic regression model. The predictive model also adjusted for both sampling weights and jackknife replicate weights for each wave of the outcome measures. The estimated parameters were then applied to the treatment group to estimate the probabilities of choosing these arrangements for Head Start participants if they had been assigned to the control group. These probabilities are referred to as principal scores since they are used to stratify the population into mutually exclusive subgroups (i.e., principal strata) based on theoretical pre-treatment variables (Frangakis & Rubin, 2002; Hill et al., 2003).

In the second stage, each child in the control group who received a specific child care arrangement was matched with Head Start participants in the treatment group who had the closest principal scores, using radius matching with a caliper at 0.01 (i.e., much smaller than 0.25 times a standard deviation of the predicted principal scores, as suggested by Rosenbaum & Rubin, 1985). Radius matching with a caliper allows for the use of all comparison units within the maximum distance of the caliper where best matches can be made (Dehejia & Wahba, 2002; Neidell & Waldfogel, 2009). In addition, a common support option was used in the matching to limit children with the specific child care arrangement to those whose principal scores had overlap with those of Head Start participants in the treatment group. The random assignment of the HSIS ensured that children in the treatment and control groups overall were similar at baseline, which made it possible to find “matches” in the treatment group for children in the control group who had specific child care arrangements (Hill et al., 2002, 2003; Zhai et al., 2010). Balance tests were conducted to ensure that after matching the covariates of children in the matched samples were well-balanced (Dehejia & Wahba, 2002).

In the third stage, the effects of Head Start were estimated by the regression-adjusted differences in outcomes between children who received specific child care arrangements in the control group and matched Head Start participants in the treatment group. Regression adjustment after matching or random assignment takes into account the effects of covariates on outcomes rather than attributing children’s differences in outcomes only to their participation in Head Start and thus can further reduce potential bias (Hill et al., 2003; USDHHS, 2005, 2010, 2012). OLS regressions were conducted including the same covariates as in the above ITT analyses, using sampling weights and jackknife replicate

weights for each wave of the outcome measures multiplied, respectively, by the weights generated from the principal score matching process.

To test the robustness of the findings of OLS regressions without and with principal score matching, we further employed an inverse probability weighting (IPW) approach. IPW aims to reweight children in the comparison groups to make them representative of the population of interest and can lead to an efficient estimate of the average treatment effect (Austin, 2011; Hirano, Imbens, & Ridder, 2003). Compared to matching approaches, IPW may be more sensitive to model specifications and sometimes remove less imbalance between comparison groups (Austin, 2011; Rubin, 2004). In addition, IPW may bias the estimated standard errors downward and may also have inaccurate or unstable weights for observations with quite high or low probabilities (Austin, 2011; Freedman & Berk, 2008). Therefore, IPW may result in biased estimates of treatment effects.

To use IPW in the investigation of the treatment effects of Head Start contingent on a specific child care arrangement (i.e., what care arrangement children would receive in the absence of Head Start), the principal scores (i.e.,  $P$ ) estimated in the first stage above were used to calculate the weights for Head Start participants in the treatment group [ $P/(1 - P)$ ] and assigned a weight of 1 for children in the control group who received the specific child care arrangement. Similar to the principal score matching, sampling and jackknife replicate weights for each wave of the outcome measures, multiplied respectively by the weights calculated in IPW, were then used in the OLS regressions to estimate the effects of Head Start compared to the specific child care arrangements in the control group.

In the analyses, a large number of individual statistical tests were conducted, which increased the probability that some findings might be statistically significant by chance. To address this issue of false discovery, a Benjamini-Hochberg (1995) test was conducted to limit the false discovery rate in the outcome measures used for a given age cohort in a given year to no more than 10 percent, following the procedure adopted in the analyses for the HSIS reports (USDHHS, 2010, 2012). In this test, the original  $p$ -values for the individual impact estimates were ranked from 1 to  $m$ , where  $m$  was the total number of effects estimated for the outcome measures in a given cohort and a given year. Each  $p$ -value was then compared to a calculated value equal to the value of its rank position in the ordering multiplied by 0.05 and divided by  $m$ . If a particular estimate was smaller than this calculated value, it was declared to pass the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.

## Results

### Descriptive Statistics

Table 2 presents the distribution of child care arrangements in the full sample of analysis in the Head Start year of spring 2003 ( $n = 3,778$ ) and by treatment status for the 3- and 4-year-old cohorts separately. Not surprisingly, the majority of children (over third quarters) in the Head Start assigned treatment group participated in Head Start (85% in the 3-year-old cohort and 78% in the 4-year-old cohort). In the control group, the most frequent child care arrangement was parental care (about 40% of children in both cohorts), followed by other

center-based care (25% in the 3-year-old cohort and 37% in the 4-year-old cohort). In addition, a considerable proportion of children in the control group also managed to attend a Head Start program (16% in the 3-year-old cohort and 13% in the 4-year-old cohort).

Table 3 shows the descriptive statistics of child and family covariates in the Head Start year of spring 2003 by the treatment status and types of child care arrangements for 3- and 4-year-old cohorts separately (adjusted by sampling weights and jackknife replicate weights). The mean differences between Head Start participants in the treatment group and children in the control group with different care arrangements were tested using regression models. The discussion below focuses on results that were statistically significant at  $p < 0.05$ .

As presented in Table 3, since all children in the HSIS were eligible for Head Start, children who were randomly assigned to the treatment and control groups were similar, in contrast to the dramatic differences between Head Start and non-Head Start children usually found in observational studies (e.g., R. Lee et al., 2014; V. Lee et al., 1988; Zhai et al., 2011). However, differences did emerge when comparing Head Start participants in the treatment group to children in the control group who attended specific types of care arrangements. For example, in the 3-year-old cohort, compared to Head Start participants in the treatment group, children in the control group who received parental care or relative non-relative care were less likely to have special needs and more likely have younger mothers and live with both biological parents. In the 4-year-old cohort, compared to Head Start participants in the treatment group, children in the control group who had parental care tended to be Black and have low or no household risk. Relatively fewer differences were found between Head Start participants in the treatment group and children in the control group who attended Head Start or other center-based care.

Appendix Table 1 shows the descriptive results of covariates after principal score matching in spring 2003 as an example of balance tests, in which no statistically significant differences were detected.

### Head Start and Children's Cognitive Outcomes

Table 4 presents the effects of Head Start on children's cognitive and behavioral outcomes using principal score matching (PSM) by cohort and year of data collection. Effect sizes (denoted below as  $d$ ) reported in the table were calculated from the regression coefficients divided by the standard deviations of the measures in the control group after using sampling weights (USDHHS, 2005, 2010, 2012). The raw regression coefficients with associated standard errors from PSM and those from OLS and IPW models are shown in Table 5. The bold results in Table 4 suggest that the significant findings passed the Benjamini-Hochberg tests for multiple comparisons. We adopted the typology of evidence used in the HSIS reports (USDHHS, 2010, 2012) based on the significance levels of individual tests and the Benjamini-Hochberg tests for multiple comparisons with a 10 percent false discovery rate. Strong evidence means that the estimated impact was significant at  $p < 0.05$  and held up after adjusting for false discovery. Moderate evidence indicates that the estimated impact was significant at  $p < 0.05$  but did not hold up after adjusting for false discovery. Suggestive evidence indicates that the estimated impact was marginally significant at  $p < 0.10$  and may or may not hold up after adjusting for false discovery. Given the large number of results

presented in Table 4 and other tables, the discussions below focus on results that met the strong or moderate evidence standard (results that provided only suggestive evidence are shown in the tables but not discussed).

The ITT and TOT estimates reported in Table 5 were quite close, if not identical, to those presented in the HSIS report (USDHHS, 2010), with negligible differences existing possibly due to the exclusions of Puerto Rico data in the restricted-use data and children with missing data on child care arrangements.<sup>4</sup> In the comparisons of Head Start in the treatment group to the specific child care arrangements in the control group, overall, the results in Table 5 from IPW and PSM models were consistent in terms of statistical significance and magnitude in most models. Since the PSM models may better address the issue of selection into alternative child care arrangements, the discussions below focus on the results of these models.

The results shown in Table 4 indicate that the effects of Head Start varied substantially by comparison group and that the size of significant effects when doing the more targeted comparisons with specific care arrangements was consistently larger than the ITT estimates. Specifically, compared to parental care, Head Start had significant effects on cognitive outcomes in spring of the Head Start year in 2003. In the 3-year-old cohort, strong evidence suggests that Head Start was associated with improvements in PPVT ( $d = 0.30$ ), WJ-III Word ( $d = 0.51$ ), and WJ-III Applied Problems ( $d = 0.33$ ). In the 4-year-old cohort in spring of the Head Start year, strong evidence shows that compared to parental care, Head Start was associated with higher scores on PPVT ( $d = 0.30$ ), WJ-III Word ( $d = 0.46$ ), and WJ-III Applied Problems ( $d = 0.36$ ).

When compared to parental care, Head Start also showed some sustained effects on cognitive outcomes. In the 3-year-old cohort, there was strong evidence that Head Start had positive effects on WJ-III Word ( $d = 0.30$ ) through age 4. There was also moderate evidence that, compared to parental care, the effects of Head Start on WJ-III Applied Problems were sustained through age 4 ( $d = 0.16$ ) and first grade ( $d = 0.16$ ). In the 4-year-old cohort, compared to parental care, Head Start was associated with higher scores on PPVT through kindergarten ( $d = 0.19$ , moderate evidence) and first grade ( $d = 0.25$ , strong evidence), on WJ-III Word through first grade ( $d = 0.23$ , moderate evidence), and on WJ-III Applied Problems through kindergarten ( $d = 0.25$ , moderate evidence) and first grade ( $d = 0.30$ , strong evidence).

The findings on the effects of Head Start on cognitive outcomes compared to relative or non-relative care largely paralleled those from the comparison of Head Start to parental care. As shown in Table 4, in the 3-year-old cohort, Head Start showed significant initial effects on PPVT ( $d = 0.19$ ), WJ-III Word ( $d = 0.52$ ), and WJ-III Applied Problems ( $d = 0.30$ ) in spring of the Head Start year in 2003. In the 4-year-old cohort, compared to relative/non-relative care, strong evidence shows that Head Start was associated with increased scores on WJ-III Word ( $d = 0.72$ ) in spring of the Head Start year.

---

<sup>4</sup>Since the TOT estimates were calculated as the ITT estimates divided by  $(1 - n - c)$ , where  $n$  is the no-show rate and  $c$  is the crossover rate, the statistical significance levels were identical for these two sets of estimates.

There were also some sustained effects of Head Start on cognitive outcomes when compared to relative/non-relative care. In the 3-year-old cohort, there was moderate evidence that the effects of Head Start on PPVT ( $d = 0.19$ ) and WJ-III Word ( $d = 0.23$ ) persisted through age 4. No sustained effects were found in the 4-year-old cohort.

In contrast, we found no strong evidence that Head Start had significant effects on cognitive outcomes compared to other center-based care or Head Start in the control group. In the 3-year-old cohort, there was moderate evidence that Head Start was related to increased PPVT in spring of the Head Start year ( $d = 0.18$ ) compared to other center-based care and to increased PPVT in first grade ( $d = 0.21$ ) compared to Head Start in the control group. In the 4-year-old cohort, there were no significant differences in cognitive outcomes when Head Start in the treatment group was compared to other center-based care or Head Start in the control group.

### Head Start and Parent-reported Social-Behavioral Outcomes

The lower panel of Table 4 shows the effects of Head Start on parent-reported social and behavioral outcomes by cohort and by year of data collection. Similar to the findings on cognitive outcomes, the strongest results were from the comparisons of Head Start to parental care and relative/non-relative care. Many of the significant effects of Head Start in both cohorts on parent-reported aggressive and hyperactive behavior compared to specific alternative care arrangements were not evident at all in the general ITT models.

In both the 3- and 4-year-old cohorts, when compared to parental care, Head Start had significant initial effects on the reduction of parent-reported Hyperactive Behavior ( $d = -0.35$  in the 3-year-old cohort, strong evidence, and  $d = -0.19$  in the 4-year-old cohort, moderate evidence) in spring of the Head Start year in 2003. In the 3-year-old cohort, the effects of Head Start on Hyperactive Behavior were also sustained through age 4 ( $d = -0.16$ , moderate evidence) and kindergarten ( $d = -0.16$ , moderate evidence). In addition, in the 3-year-old cohort, compared to parental care, Head Start was linked to reduced Aggressive Behavior through age 4 ( $d = -0.15$ , moderate evidence) and kindergarten ( $d = -0.19$ , strong evidence).

Compared to relative/non-relative care, strong evidence suggests that in the 3-year-old cohort Head Start showed significant initial effects on the reduction of Aggressive Behavior ( $d = -0.23$ ) and Hyperactive Behavior ( $d = -0.26$ ) in spring of the Head Start year in 2003. Similarly, in the 4-year-old cohort, strong evidence also shows that Head Start was also associated with reduced Aggressive Behavior ( $d = -0.44$ ) in spring of the Head Start year.

There were some sustained effects of Head Start on parent-reported behavioral outcomes compared to relative/non-relative care. In the 3-year-old cohort, Head Start was related to lower scores on Aggressive Behavior through age 4 ( $d = -0.26$ , moderate evidence) and first grade ( $d = -0.26$ , moderate evidence) and on Hyperactive Behavior through kindergarten ( $d = -0.27$ , strong evidence) and first grade ( $d = -0.25$ , moderate evidence). In the 4-year-old cohort, compared to relative/non-relative care, Head Start was related to decreased Aggressive Behavior through kindergarten ( $d = -0.31$ , strong evidence) and first grade ( $d =$

–0.38), while there was moderate evidence that Head Start was associated with reduced Hyperactive Behavior through first grade ( $d = -0.40$ ).

Similar to the findings on cognitive outcomes, we did not find strong evidence that Head Start had significant effects on parent-reported behavioral outcomes compared to other center-based care or Head Start in the control group, except for the effects on Social Skills and Positive Approaches to Learning in kindergarten in the 3-year-old cohort ( $d = 0.34$ ). In the 3-year-old cohort, compared to other center-based care, moderate evidence suggests that Head Start was associated with reduced Aggressive Behavior at age 4 ( $d = -0.25$ ) and Hyperactive Behavior in kindergarten ( $d = -0.19$ ).

## Discussion

We used data from the HSIS, the only large-scale randomized experiment in Head Start history, to examine a question the random assignment study could not address – whether and how the impact of Head Start varied contingent on the alternative child care arrangements to which it was compared. Using a principal score matching approach to address the issue of selection into child care arrangements in the control group, we found that the effects of Head Start varied substantially depending on the alternative care arrangement, with the strongest and most lasting effects evident when Head Start was compared to parental care or relative/non-relative care. By comparing Head Start children in the treatment group with matched comparisons in several types of alternative care arrangements, this approach to analysis helped uncover significant effects of Head Start that the general ITT findings did not demonstrate and provided a different and more nuanced take-home story than “Head Start/preschool does not work.”

It should be noted that principal score matching, like propensity score matching and regressions in general, assumes ignorability or selection on observables, but not unobservables. In other words, it relies on the assumption that all confounding covariates related to treatment status are observed (Dehejia & Wahba, 2002; Hill et al., 2003; Rosenbaum & Rubin, 1985). If any important variables unrelated to the covariates that were included in the models are omitted, the estimates of Head Start effects could possibly be biased. For example, parents’ educational expectation and motivation have been found positively associated with children’s early academic and social-behavioral development (Fan & Chen, 2001; Galindo & Sheldon, 2012; Hong & Ho, 2005; Kim, Sheridan, Kwon, & Koziol, 2013). These factors may also affect parents’ care arrangements for their children and thus should be investigated in future research. Moreover, due to the relatively small sample sizes in the comparisons of some specific child care arrangements by cohort, the estimates of Head Start effects may not be precise and thus caution should be taken in interpreting and generalizing the findings. In addition, we used the focal child care arrangements defined in the HSIS reports to be consistent with the analyses in these reports; this variable also had minimal missing data. A considerable proportion of children in the sample had missing data on the time they spent in specific child care settings (e.g., 37% with missing data on the hours in the setting that they spent most of the time from Monday through Friday). Therefore, given the large number of models conducted, we did not further explore the effects of child care arrangements with various hours or multiple care

arrangements, which may introduce some systematic bias into the data and thus remain as important topics to be investigated in the future.

In spite of these limitations, the findings in this study on how the effects of Head Start vary depending on the specific child care arrangements in the control group provide important information as to the children for whom the benefits of Head Start are likely to be greatest. In particular, the findings on the benefits of Head Start compared to parental care and relative/non-relative care are important since the most common arrangement in the HSIS for eligible children whose parents applied for but did not get into Head Start was exclusive parental care (41% in the 3-year-old cohort and 38% in the 4-year-old cohort, as shown in Table 2), and together with children in relative/non-relative care, they account for half or more of the children who were not granted access to Head Start (59% in the 3-year-old cohort and 50% in the 4-year-old cohort). These children, especially those in the 3-year-old cohort, may also benefit from Head Start through reductions of problem behaviors. Therefore, if the goal of Head Start is to improve children's school readiness in terms of cognitive and behavioral outcomes, an implication of this research is that policymakers should modify intake procedures for Head Start programs to ensure they reach disadvantaged children who otherwise would be likely to stay home with their parents or receive care from a relative or non-relative. To do this, one approach would be to identify communities that have relatively large proportions of low-income families but do not currently have many preschool educational facilities. Where a large discrepancy exists, that would signal a need to develop and fund new centers (via possibly enhanced federal and state funding). In addition, within communities, it would also be possible to identify Head Start and other community centers that have the capacity to serve more children. Within Head Start programs, it might also be possible to implement better and more rapid recording of take-up, to track, for example, what proportion of children who are enrolled in the fall but leave the program during the year. These too are slots that could be filled. Once open slots are identified (within communities and within Head Start programs), outreach could be undertaken to find eligible families who are not being served, akin to barefoot doctors or public health initiatives to locate children and families in need of services, which have been often used in developing countries but sometimes in the U.S. as well (e.g., a review by Lehmann & Sanders, 2007, for the World Health Organization; also see research and reviews by Bangdiwala et al., 2011; McCormick et al., 1989; Nxumalo, Goudge, & Thomas, 2013; V. Lee et al., 1988; van Ginneken, Lewin, & Berridge, 2010).

In contrast, our results indicate that Head Start generally compared well to other center-based care. Overall, we found only a few differences between Head Start and other center-based care in the 3-year-old cohort and no differences in the 4-year-old cohort. This finding may not be surprising given the characteristics of the two types of programs based on data provided by the HSIS. On the one hand, Head Start in the HSIS treatment group had higher overall quality (with quality composite scores of 0.71 in the 3-year-old cohort and 0.74 in the 4-year-old cohort)<sup>5</sup> than other center-based care in the control group (with quality

---

<sup>5</sup>The overall quality composite score was created using 12 variables from observation ratings, activities provided in the setting, teacher qualifications and experiences, parent involvement, home visits, and program services collected in spring 2003 (USDHHS, 2010).

composite scores of 0.55 in the 3-year-old cohort and 0.57 in the 4-year-old cohort) and had more directors with a Bachelor's degree (71% vs. 48% of directors in other center-based care in the 3-year-old cohort, but not in the 4-year-old cohort). On the other hand, other center-based care in the control group had more teachers with a Bachelor's degree (50% in the 3-year-old cohort and 48% in the 4-year-old cohort) than Head Start in the treatment group (34% in the 3-year-old cohort and 32% in the 4-year-old cohort) and a lower turnover rate in the 3-year-old cohort (14% for new lead teachers vs. 18% in Head Start). Therefore, while other center-based care may have overall lower quality than Head Start, children who attend other center-based care may be compensated by more teachers with a Bachelor's degree and staying with same teachers. This raises the possibility that there is still much room to improve the quality of Head Start programs and children's experiences through measures such as increasing the rates of teachers and directors with a Bachelor's degree and reducing teachers' turnover rate. It should be noted that the definition and measures of early childhood education classroom quality have evolved significantly in recent years. There is now much more focus on process quality, especially the quality of teacher-child interactions, which has been found to be a strong predictor of children's sustainable gains in language, literacy, mathematics, and social skills (Burchinal, Vandergrift, Pianta, & Andrew, 2010; Phillips & Lowenstein, 2011; Yoshikawa et al., 2013). Since the element of process quality was not well represented in the HSIS measures, we could not compare the process quality between Head Start and other center-based care, which could be addressed in future research. Further analysis is also needed to look more closely at other outcomes that Head Start programs may address whereas other center-based programs may not (e.g., health, insurance coverage, and dental care).

In addition, we found almost no benefits of Head Start in social skills and approaches to learning (with one exception in kindergarten in the 3-year-old cohort). This finding is somewhat surprising, given Head Start's emphasis on the whole child. The approaches to learning skills are also one of the five essential domains (along with language and literacy, cognition and general knowledge, physical development and health, and social and emotional development) in the school readiness goals that all Head Start agencies are required to establish and take steps to achieve (USDHHS, 2011). One possible explanation may be the limitation of the parent-reported measure of social skills and approaches to learning used in our study, which may not provide a good estimate of the construct. Future research could use teacher-reported measures of social and learning skills in the HSIS data to investigate whether Head Start has significant effects on them. Moreover, we found that some effects of Head Start were not sustained after the Head Start year (e.g., on PPVT when compared to parental care and on WJ-III Applied Problems compared to relative/non-relative care in the 3-year-old cohort), while other effects that were sustained tended to go down over time. Further scientific inquiries are needed to better understand the follow-up contexts that children are exposed to after the Head Start year (e.g., a second year of Head Start for the 3-year-old cohort, and elementary school classrooms for both cohorts).

In conclusion, do the effects of Head Start vary depending on the alternative child care arrangements to which it is compared? The answer seems to be yes. Head Start clearly is most beneficial for children who, in the absence of the program, would have remained home with their parents or would have received informal care from a relative or non-relative, and



for these children at least some of the benefits of Head Start participation persist through first grade. Therefore a clear policy implication of this study is to ensure that Head Start, and other center-based, programs reach more children who could benefit from them.

## Acknowledgments

This research was supported by a grant from the American Educational Research Association which receives funds for its “AERA Grants Program” from the National Science Foundation under NSF Grant #DRL-0941014. The authors also thank the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) through grant R24HD058486 to the Columbia Population Research Center. The opinions and conclusions expressed herein are solely those of the authors and should not be construed as representing the opinions or policies of the granting agencies, the AERA or any agency of the Federal government.

## References

- Aber JL, Brown JL, Jones SM. Developmental trajectories toward violence in middle childhood: Course, demographic differences, and response to school-based intervention. *Developmental Psychology*. 2003; 39:324–348. [PubMed: 12661889]
- Achenbach TM, Edelbrock C, Howell CT. Empirically based assessment of the behavioral/emotional problems of 2-3-year old children. *Journal of Abnormal Child Psychology*. 1987; 15:629–650. [PubMed: 3437096]
- Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*. 2011; 46:399–424. [PubMed: 21818162]
- Bangdiwala SI, Tucker JD, Zodpey SP, Griffiths SM, Li L, Reddy KS, et al. Public health education in India and China: History, opportunities, and challenges. *Public Health Reviews*. 2011; 33:204–224.
- Barnard J, Frangakis CE, Hill LH, Rubin DB. Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in New York City. *Journal of the American Statistical Association*. 2003; 98:299–323.
- Baydar N, Brooks-Gunn J. Effects of maternal employment and child-care arrangements on preschoolers’ cognitive and behavioral outcomes: Evidence from the children of the National Longitudinal Survey of Youth. *Developmental Psychology*. 1991; 27:932–945.
- Benjamini Y, Hochberg Y. Controlling for the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)*. 1995; 57:289–300.
- Besharov DJ, Call DM. Head Start falls further behind. *New York Times*. 2009 Feb 8. (p. WK12 of the New York edition).
- Brooks-Gunn, J. Early childhood education: The likelihood of sustained effects. In: Zigler, E.; Gilliam, WS.; Barnett, WS., editors. *The pre-K debates: Current controversies and issues*. Baltimore, MD: Brookes; 2011. p. 200-206.
- Burchinal M, Vandergrift N, Pianta R, Andrew M. Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly*. 2010; 25:166–176.
- Camilli G, Vargas S, Ryan S, Barnett WS. Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record*. 2010; 112:579–620.
- Currie J. Health disparities and gaps in school readiness. *The Future of Children*. 2005; 15:117–138. [PubMed: 16130544]
- Currie J, Thomas D. Does Head Start make a difference? *American Economic Review*. 1995; 85:341–364.
- Currie J, Thomas D. Does Head Start help Hispanic children? *Journal of Public Economics*. 1999; 74:235–262.
- Dehejia R, Wahba S. Propensity score matching methods for nonexperimental causal studies. *The Review of Economics and Statistics*. 2002; 84:151–161.
- Deming D. Early childhood intervention and life-cycle skill development: Evidence from Head Start. *American Economic Journal: Applied Economics*. 2009; 1:111–134.

- Dunn, LM.; Dunn, LL.; Dunn, DM. Peabody Picture and Vocabulary Test, Third Edition (PPVT). Circle Pines, MN: American Guidance Service; 1997.
- Fan X, Chen M. Parental involvement and students' academic achievement: A meta-analysis. *Educational Psychology Review*. 2001; 13:1–22.
- Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics*. 2002; 58:21–29. [PubMed: 11890317]
- Freedman DA, Berk RA. Weighting regressions by propensity scores. *Evaluation Review*. 2008; 32:392–409. [PubMed: 18591709]
- Galindo C, Sheldon SB. School and home connections and children's kindergarten achievement gains: The mediating role of family involvement. *Early Childhood Research Quarterly*. 2012; 27:90–103.
- Garces E, Thomas D, Currie J. Longer term effects of Head Start. *The American Economic Review*. 2002; 92:999–1012.
- Gormley W Jr. The effects of Oklahoma's pre-K program on Hispanic children. *Social Science Quarterly*. 2008; 89:916–936.
- Hill LJ, Brooks-Gunn J, Waldfogel J. Sustained effects of high participation in an early intervention for low-birth-weight premature infants. *Developmental Psychology*. 2003; 39:730–744. [PubMed: 12859126]
- Hill LJ, Waldfogel J, Brooks-Gunn J. Differential effects of high-quality child care. *Journal of Policy Analysis and Management*. 2002; 21:601–627.
- Hirano K, Imbens GW, Ridder G. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*. 2003; 71:1161–1189.
- Hong S, Ho HZ. Direct and indirect longitudinal effects of parental involvement on student achievement: Second-order latent growth modeling across ethnic groups. *Journal of Educational Psychology*. 2005; 97:32–42.
- Karoly, LA.; Kilburn, MR.; Cannon, J. Early childhood interventions: Proven results, future promise. Santa Monica, CA: RAND Corp; 2005.
- Kim EM, Sheridan SM, Kwon K, Koziol N. Parent beliefs and children's social-behavioral functioning: The mediating role of parent-teacher relationships. *Journal of School Psychology*. 2013; 51:175–185. [PubMed: 23481083]
- Lee R, Zhai F, Brooks-Gunn J, Han WJ, Waldfogel J. Head Start participation and school readiness: Evidence from the Early Childhood Longitudinal Study-Birth Cohort. *Developmental Psychology*. 2014; 50:202–215. [PubMed: 23527496]
- Lee VE, Brooks-Gunn J, Schnur E. Does Head Start work? A 1-year follow-up comparison of disadvantaged children attending Head Start, no preschool, and other preschool programs. *Developmental Psychology*. 1988; 24:210–222.
- Lehmann, U.; Sanders, D. Community health workers: What do we know about them? The state of the evidence on programmes, activities, costs and impact on health outcomes of using community health workers. Geneva, Switzerland: World Health Organization; 2007.
- Love JM, Chazan-Cohen R, Raikes H, Brooks-Gunn J. What makes a difference: Early Head Start evaluation findings in a developmental context. *Monographs of the Society for Research in Child Development*. 2013
- Ludwig J, Miller DL. Does Head Start improve children's life chances? Evidence from a regression discontinuity design. *Quarterly Journal of Economics*. 2007; 122:159–208.
- Magnuson KA, Ruhm C, Waldfogel J. Does prekindergarten improve school preparation and performance? *Economics of Education Review*. 2007; 26:33–51.
- Magnuson KA, Waldfogel J. Early childhood care and education: Effects on ethnic and racial gaps in school readiness. *The Future of Children*. 2005; 15:169–196. [PubMed: 16130546]
- McCormick MC, Brooks-Gunn J, Shorter T, Holmes JH, Wallace CY, Heagarty MC. Outreach as case finding: Its effect on enrollment in prenatal care. *Medical Care*. 1989; 27:103–111. [PubMed: 2918763]
- Neidell M, Waldfogel J. Program participation of immigrant children: Evidence from the local availability of Head Start. *Economics of Education Review*. 2009; 28:704–715.

- NICHD Early Child Care Research Network. Early child care and children's development in the primary grades: Follow-up results from the NICHD Study of Early Child Care. *American Educational Research Journal*. 2005; 42:537–570.
- Duncan GJ. NICHD Early Child Care Research Network. Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development*. 2003; 74:1454–1475. [PubMed: 14552408]
- Nisbett RE. Education is all in your mind. *New York Times*. 2009 Feb 8. (p. WK12 of the New York edition).
- Nxumalo N, Goudge J, Thomas L. Outreach services to improve access to health care in South Africa: Lessons from three community health worker programmes. *Global Health Action*. 2013; 6:219–226.
- Phillips DA, Lowenstein AE. Early care, education, and child development. *Annual Review of Psychology*. 2011; 62:483–500.
- Raver CC, Jones SM, Li-Grining CP, Zhai F, Metzger MW, Solomon B. Targeting children's behavior problems in preschool classrooms: A cluster-randomized controlled trial. *Journal of Consulting and Clinical Psychology*. 2009; 77:302–316. [PubMed: 19309189]
- Reid MJ, Webster-Stratton C, Baydar N. Halting the development of conduct problems in Head Start children: The effects of parent training. *Journal of Clinical Child and Adolescent Psychology*. 2004; 33:279–291. [PubMed: 15136193]
- Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Journal of the American Statistician*. 1985; 39:33–38.
- Rubin DB. On principles for modeling propensity scores in medical research. *Pharmacoepidemiology Drug Safety*. 2004; 13:855–857. [PubMed: 15386710]
- Smolensky, E.; Gootman, JA. *Working families and growing kids: Caring for children and adolescents*. Washington, DC: The National Academies Press; 2003.
- Styfco, S.; Zigler, E. *The Head Start debates*. Baltimore, MD: Brookes; 2004.
- Tolan P, Gorman-Smith D, Henry D. Supporting families in a high-risk setting: Proximal effects of the SAFE Children preventive intervention. *Journal of Consulting and Clinical Psychology*. 2004; 72:855–869. [PubMed: 15482043]
- U.S. Department of Health and Human Services. *Head Start Impact Study: First year findings*. Washington, DC: Author; 2005. USDHHS
- U.S. Department of Health and Human Services. *Head Start Impact Study: Final report*. Washington, DC: Author; 2010. USDHHS
- U.S. Department of Health and Human Services. *Head Start approach to school readiness*. Washington, DC: Author; 2011. USDHHS
- U.S. Department of Health and Human Services. *Third grade follow-up to the Head Start Impact Study: Final report*. Washington, DC: Author; 2012. USDHHS
- van Ginneken N, Lewin S, Berridge V. The emergence of community health worker programmes in the late apartheid era in South Africa: An historical analysis. *Social Science & Medicine*. 2010; 71:1110–1118. [PubMed: 20638169]
- Waldfoegel, J. *What children need*. Cambridge, MA: Harvard University Press; 2006.
- Woodcock, RW.; McGrew, KS.; Mather, N. *Woodcock-Johnson III*. Itasca, IL: Riverside; 2001.
- Yoshikawa, H.; Weiland, C.; Brooks-Gunn, J.; Burchinal, MR.; Espinosa, LM.; Gormley, WT.; Magnuson, KA.; Phillips, D.; Zaslow, MJ. *Investing in our future: The evidence base on preschool education*. Ann Arbor, MI: Society for Research in Child Development; 2013. New York: Foundation for Child Development
- Zhai F, Brooks-Gunn J, Waldfoegel J. Head Start and urban children's school readiness: A birth cohort study in 18 cities. *Developmental Psychology*. 2011; 47:134–152. [PubMed: 21244155]
- Zhai F, Raver CC, Jones S, Li-Grining C, Pressler E, Gao Q. Dosage effects on school readiness: Evidence from a randomized classroom-based intervention. *Social Service Review*. 2010; 84:615–654. [PubMed: 21488322]

## Appendix

Appendix Table 1

Balance check of covariates after principal score matching in spring 2003

	HS in Treatment vs. HS in Control		HS in Treatment vs. Other Center-based		HS in Treatment vs. Relative/Non-relative		HS in Treatment vs. Parental	
	a	b	a	b	a	b	a	b
<u>3-year-old cohort</u>								
Girl	0.50	0.48	0.51	0.51	0.51	0.49	0.56	0.55
Age (weeks)	214.48	214.93	213.64	214.60	214.40	213.56	213.98	214.24
Race								
White/Other	0.32	0.38	0.28	0.31	0.25	0.26	0.37	0.37
Black	0.38	0.35	0.41	0.38	0.40	0.41	0.25	0.26
Hispanic	0.30	0.26	0.31	0.31	0.35	0.33	0.38	0.37
Test language in Spanish	0.20	0.19	0.17	0.18	0.15	0.13	0.27	0.26
Special needs of child	0.16	0.18	0.15	0.15	0.07	0.05	0.10	0.09
Mother age	29.24	28.62	29.98	29.65	28.49	27.92	29.54	29.55
Both bio-parents live with	0.51	0.54	0.49	0.52	0.37	0.36	0.54	0.57
Mother immigrant	0.18	0.20	0.15	0.14	0.13	0.11	0.18	0.18
Primary home language	0.75	0.71	0.73	0.70	0.80	0.81	0.69	0.69
Household risk index								
Low/no	0.74	0.76	0.79	0.82	0.84	0.87	0.78	0.77
Medium	0.23	0.22	0.12	0.11	0.11	0.11	0.16	0.18
High	0.03	0.02	0.09	0.07	0.04	0.03	0.06	0.06
Urbanicity	0.83	0.82	0.83	0.85	0.75	0.76	0.77	0.76
<u>4-year-old cohort</u>								
Girl	0.37	0.34	0.51	0.49	0.50	0.51	0.49	0.51
Age (weeks)	262.13	265.07	262.96	264.04	261.64	263.31	261.18	262.34
Race								
White/Other	0.22	0.21	0.34	0.36	0.33	0.35	0.36	0.34
Black	0.27	0.25	0.29	0.29	0.25	0.24	0.16	0.15
Hispanic	0.51	0.54	0.37	0.34	0.41	0.40	0.48	0.51
Test language in Spanish	0.36	0.39	0.29	0.25	0.32	0.30	0.33	0.34
Special needs of child	0.14	0.15	0.15	0.13	0.12	0.10	0.10	0.09
Mother age	29.67	29.67	29.13	29.07	29.12	29.05	29.35	29.33
Both bio-parents live with	0.56	0.53	0.50	0.50	0.52	0.50	0.53	0.57
Mother immigrant	0.31	0.28	0.23	0.20	0.26	0.22	0.25	0.26
Primary home language	0.54	0.57	0.69	0.73	0.64	0.66	0.63	0.64
Household risk index								
Low/no	0.74	0.76	0.76	0.79	0.77	0.80	0.76	0.78
Medium	0.23	0.21	0.19	0.15	0.17	0.13	0.19	0.16
High	0.03	0.03	0.05	0.06	0.06	0.07	0.05	0.06
Urbanicity	0.90	0.90	0.83	0.81	0.83	0.85	0.82	0.84

Notes: means presented in table were computed in matched samples adjusted by sampling weights and jackknife replicate weights generated in principal score matching models; regression models (OLS for continuous measures and logistic

regressions for binary measures) with sampling weights and jackknife replicate weights from principal score matching models were used to test mean differences between matched Head Start participants in the treatment group (a) and children in the control group with different care arrangements (b), including Head Start (HS), other center-based care, relative/non-relative care, and parental care; no statistically significant differences were detected.

**Table 1**

Data collection and sample size in analysis

	Fall 2002	Spring 2003	Spring 2004	Spring 2005	Spring 2006
3-year-old cohort	Baseline, Head Start (n = 2,119)	Head Start (n = 2,111)	4-year-old, Head Start (n = 2,005)	Kindergarten (n = 1,907)	First grade (n = 1,867)
4-year-old cohort	Baseline, Head Start (n = 1,671)	Head Start (n = 1,667)	Kindergarten (n = 1,505)	First grade (n = 1,502)	No data collected
Full analysis sample	3,790	3,778	3,510	3,409	1,867

Notes: the number of observations presented in table was based on non-missing data of child care arrangements in spring 2003 and outcome variables by year and by cohort.

**Table 2**

Distribution of child care arrangements in analysis sample of spring 2003 (%)

	Full Sample	3-year-old cohort			4-year-old cohort		
		All	Treatment	Control	All	Treatment	Control
Head Start	56.62	59.02	84.70	16.18	53.57	78.01	13.34
Other center-based care	16.60	13.03	5.83	25.03	21.11	11.48	36.98
Relative/non-relative care	7.12	8.15	2.12	18.20	5.82	2.41	11.43
Parental care	19.66	19.80	7.35	40.58	19.50	8.10	38.25
Sample size (n)	3,778	2,111	1,320	791	1,667	1,037	630

Notes: percentage presented in table was based on non-missing data of child care arrangements and outcome variables in spring 2003.

**Table 3**

Descriptive statistics of child and family covariates by child care arrangements

	3-year-old cohort				4-year-old cohort					
	HS in Treatment	HS	Oth. Ctr.	Rel./Non-rel.	Parent.	HS in Treatment	HS	Oth. Ctr.	Rel./Non-rel. Parent.	
Girl	0.54	0.47	0.51	0.49	0.55	0.50	0.34*	0.49	0.54	0.52
Age (weeks)	214.17	214.93	214.59	213.88	215.23	260.77	265.07	263.70+	262.66	262.37
Race										
White/Other	0.31	0.38	0.30	0.26	0.38	0.33	0.21+	0.36	0.42	0.34
Black	0.36	0.35	0.38	0.40	0.25**	0.25	0.25	0.31	0.28	0.15*
Hispanic	0.33	0.26	0.32	0.35	0.37	0.42	0.54	0.34+	0.30	0.51+
Test language in Spanish	0.20	0.19	0.18	0.13+	0.27*	0.32	0.39	0.24+	0.26	0.34
Special needs of child	0.14	0.18	0.16	0.04*	0.08*	0.13	0.15	0.12	0.09	0.09
Mother age	29.56	28.62	29.84	27.31**	28.49*	29.27	29.67	28.86	30.11	29.26
Bio-parents living with	0.49	0.54	0.52	0.35*	0.57*	0.51	0.53	0.50	0.44	0.57
Mother immigrant	0.16	0.21	0.15	0.08*	0.18	0.25	0.27	0.19	0.19	0.26
Primary home language	0.75	0.71	0.70	0.82+	0.69+	0.64	0.57	0.74*	0.70	0.64
Household risk index										
Low/no	0.76	0.76	0.82+	0.87**	0.76	0.72	0.76	0.77	0.85+	0.81*
Medium	0.17	0.22	0.11*	0.10*	0.18	0.21	0.21	0.15+	0.13	0.11*
High	0.08	0.02**	0.07	0.03*	0.06	0.07	0.03+	0.08	0.03	0.07
Urbanicity	0.79	0.82	0.85	0.74	0.76	0.83	0.90	0.82	0.82	0.84

Notes: means presented in table were adjusted by sampling weights and jackknife replicate weights; regression models (OLS for continuous measures and logistic regressions for binary measures) with sampling weights and jackknife replicate weights were used to test the mean differences between Head Start participants in the treatment group and children in the control group with different care arrangements, including Head Start (HS), other center-based care (Oth. Ctr.), relative/non-relative care (Rel./Non-rel.), and parental care (Parent.), with significance levels, if statistically significant, being indicated on child care arrangements in the control group:

\*\* p<0.01,  
\* p<0.05,



+ p<0.10.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 4

## Head Start and children's outcomes

	Assignment Status (ITT)	HS in Treatment vs. Specific Care in Control			
		Parental	Relative/ Non-relative	Other Center-based	
Cognitive Outcomes					
3-year-old cohort					
	2003 Head Start	<b>0.19</b> **	<b>0.30</b> **	<b>0.19</b> *	0.19 <sup>+</sup>
PPVT	2004 Age 4	0.06	0.03	0.19*	0.01
	2005 Kindergarten	0.01	0.02	0.02	0.03
	2006 First Grade	0.08	0.00	0.04	0.12
					0.21*
4-year-old cohort					
	2003 Head Start	<b>0.26</b> **	<b>0.51</b> **	<b>0.52</b> **	0.09
WJ-III Word	2004 Age 4	0.11 <sup>+</sup>	<b>0.30</b> **	0.23*	0.03
	2005 Kindergarten	0.00	0.00	0.01	0.09
	2006 First Grade	0.01	0.05	-0.14	0.03
					0.18
	2003 Head Start	<b>0.15</b> **	<b>0.33</b> **	<b>0.30</b> **	-0.03
WJ-III Applied	2004 Age 4	0.05	0.16*	0.03	0.02
	2005 Kindergarten	-0.03	0.03	-0.12	0.05
	2006 First Grade	0.07	0.16*	-0.16	0.04
					0.15
4-year-old cohort					
	2003 Head Start	<b>0.15</b> **	<b>0.30</b> **	0.05	0.07
PPVT	2004 Kindergarten	0.08	0.19*	-0.02	-0.05
	2005 First Grade	0.13*	<b>0.25</b> **	0.00	0.04
					0.22
	2003 Head Start	<b>0.24</b> **	<b>0.46</b> **	<b>0.72</b> **	-0.02
WJ-III Word	2004 Kindergarten	0.01	0.08	-0.02	-0.05
	2005 First Grade	0.04	0.23*	-0.11	-0.02
					-0.16
WJ-III Applied	2003 Head Start	<b>0.14</b> *	<b>0.36</b> **	0.04	0.01
					0.08

	Assignment Status (ITT)	HS in Treatment vs. Specific Care in Control		
		Parental	Relative/ Non-relative	Other Center-based in Control
2004 Kindergarten	0.02	0.25*	0.12	-0.09
2005 First Grade	0.07	<b>0.30**</b>	0.04	0.01
Social-behavioral Outcomes				
3-year-old cohort				
2003 Head Start	0.02	-0.02	0.10	-0.02
2004 Age 4	0.09	0.14 <sup>+</sup>	0.23 <sup>+</sup>	0.02
2005 Kindergarten	0.14*	0.13 <sup>+</sup>	0.02	<b>0.34**</b>
2006 First Grade	0.04	0.03	0.11	0.04
2003 Head Start	-0.07	-0.06	<b>-0.23*</b>	-0.06
2004 Age 4	-0.06	-0.15*	-0.26*	-0.25*
2005 Kindergarten	-0.06	<b>-0.19*</b>	-0.18 <sup>+</sup>	-0.04
2006 First Grade	-0.03	-0.15 <sup>+</sup>	-0.26*	-0.03
Aggressive				
2003 Head Start	<b>-0.25**</b>	<b>-0.35**</b>	<b>-0.26*</b>	-0.18 <sup>+</sup>
2004 Age 4	-0.09	-0.16*	-0.11	-0.17 <sup>+</sup>
2005 Kindergarten	-0.14*	-0.16*	<b>-0.27**</b>	-0.19*
2006 First Grade	-0.09	-0.16 <sup>+</sup>	-0.25*	-0.09
4-year-old cohort				
2003 Head Start	-0.02	0.02	-0.15	0.02
2004 Kindergarten	0.05	-0.05	0.12	0.12
2005 First Grade	0.03	-0.03	-0.14	0.04
Social Skills & Approaches to Learning				
2003 Head Start	-0.10	-0.14	<b>-0.44**</b>	-0.14 <sup>+</sup>
2004 Kindergarten	-0.04	-0.01	-0.31*	0.05
2005 First Grade	-0.04	-0.10	<b>-0.38**</b>	-0.06
Aggressive				
2003 Head Start	-0.05	-0.19*	-0.07	-0.12
2004 Kindergarten	-0.05	-0.19*	-0.07	-0.12
2005 First Grade	-0.05	-0.19*	-0.07	-0.12
Hyperactive				
2003 Head Start	-0.05	-0.19*	-0.07	-0.12
2004 Kindergarten	-0.05	-0.19*	-0.07	-0.12
2005 First Grade	-0.05	-0.19*	-0.07	-0.12

	HS in Treatment vs. Specific Care in Control			
	Assignment Status (ITT)	Parental	Relative/ Non-relative	Other Center-based in Control
2004 Kindergarten	0.08	-0.07	-0.02	0.06
2005 First Grade	0.02	-0.03	-0.40*	0.01
				0.09
				0.13

Notes: effect sizes reported in table were from ITT and principal score matching models; raw regression coefficients with associated standard errors from ITT, TOT, OLS, IPW, and PSM models are presented in Table 5;

\*\*\* p<0.01,

\* p<0.05,

+ p<0.10;

**Bold results** indicate they passed the Benjamini-Hochberg tests for multiple comparisons with a 10 percent false discovery rate; strong evidence: estimate was significant at p<0.05 and held up after adjusting for false discovery; moderate evidence: estimate was significant at p<0.05 but did not hold up after adjusting for false discovery, suggestive evidence: estimate was marginally significant at p<0.10 and may or may not hold up after adjusting for false discovery.

**Table 5**

Head Start and children's outcomes: coefficients with standard errors

	HS in Treatment Group vs. Specific Care in Control Group														
	Assignment Status			Parental			Relative/Non-relative			Other Center-based			HS in Control		
	IIT	TOT	OLS	IPW	PSM	OLS	IPW	PSM	OLS	IPW	PSM	OLS	IPW	PSM	
Cognitive Outcomes															
3-year-old cohort															
	2003	6.76** (1.70)	7.17** (1.83)	8.64** (2.32)	9.15** (2.35)	10.64** (2.36)	5.92+ (3.46)	5.73+ (3.45)	6.84* (3.48)	5.57* (2.64)	5.52* (2.65)	6.35* (2.69)	7.24+ (3.91)	7.16+ (3.89)	6.91+ (3.93)
	2004	2.22 (1.83)	1.93 (1.93)	1.02 (2.52)	2.38 (2.46)	1.01 (2.52)	5.94 (3.63)	7.18* (3.38)	7.35* (3.38)	-0.72 (2.88)	-1.33 (2.87)	0.42 (3.14)	6.30 (4.16)	5.54 (4.08)	5.02 (3.88)
	2005	0.24 (1.43)	-0.47 (1.59)	0.04 (2.20)	0.11 (2.10)	0.68 (2.20)	-2.32 (2.40)	-1.22 (2.35)	0.52 (2.65)	0.26 (2.43)	-0.57 (2.55)	0.78 (2.39)	3.77 (3.14)	4.62 (3.00)	2.98 (2.79)
	2006	2.27 (1.48)	0.64 (1.66)	-0.98 (2.62)	-0.57 (2.44)	0.04 (2.47)	-0.90 (2.51)	0.55 (2.59)	1.31 (2.74)	3.75 (2.29)	2.48 (2.25)	3.52 (2.15)	9.79** (3.09)	7.16* (2.90)	6.10* (2.75)
WJ-III Word															
	2003	6.63** (1.40)	8.78** (1.46)	11.80** (1.93)	12.58** (1.78)	12.82** (1.84)	12.78** (2.20)	12.53** (2.15)	12.92** (2.24)	2.10 (2.20)	1.58 (2.28)	2.16 (2.32)	0.58 (3.18)	2.19 (3.16)	-0.08 (3.19)
	2004	3.02+ (1.60)	5.33** (1.80)	8.41** (2.44)	8.85** (2.42)	8.61** (2.45)	6.11+ (3.15)	7.65* (3.03)	6.62* (3.13)	-0.34 (2.28)	-0.81 (2.35)	0.80 (2.44)	-0.10 (3.09)	0.67 (2.99)	0.98 (2.91)
	2005	0.04 (1.88)	0.48 (2.07)	-0.42 (2.44)	0.70 (2.56)	0.00 (2.58)	-0.51 (3.73)	-0.66 (3.70)	0.31 (3.74)	1.80 (3.22)	2.04 (3.16)	3.03 (3.06)	5.07 (3.78)	5.86 (3.83)	5.85 (3.67)
	2006	0.49 (2.13)	-1.01 (2.30)	0.01 (2.99)	0.69 (3.07)	1.81 (3.00)	-7.57* (3.41)	-5.69+ (3.37)	-5.12 (3.67)	1.50 (3.53)	0.74 (3.58)	1.20 (3.64)	11.36* (4.60)	4.76 (4.48)	6.36 (4.14)
WJ-III Applied															
	2003	4.37** (1.60)	5.36** (1.69)	8.28** (2.15)	9.06** (2.27)	9.51** (2.22)	8.53** (2.69)	7.83** (2.35)	8.59** (2.74)	0.22 (2.73)	-0.76 (2.58)	-0.76 (2.59)	4.53 (3.55)	5.21 (3.49)	3.83 (3.56)
	2004	1.16 (1.25)	1.88 (1.30)	3.67* (1.71)	4.38* (1.75)	3.60* (1.71)	1.29 (2.07)	2.07 (2.06)	0.68 (2.19)	-0.69 (1.86)	-1.40 (1.91)	0.43 (2.04)	1.57 (3.36)	-0.70 (2.92)	1.39 (2.88)
	2005	-0.66 (1.17)	-0.33 (1.26)	0.02 (1.69)	1.04 (1.74)	0.65 (1.76)	-3.41+ (1.87)	-2.72 (1.81)	-2.51 (2.05)	0.69 (1.70)	0.75 (1.76)	0.98 (1.75)	3.64 (2.69)	2.90 (2.72)	2.35 (2.69)

Assignment Status	HS in Treatment Group vs. Specific Care in Control Group													
	Parental			Relative/Non-relative			Other Center-based			HS in Control				
ITT	TOT	OLS	IPW	PSM	OLS	IPW	PSM	OLS	IPW	PSM	OLS	IPW	PSM	
4-year-old cohort														
2006	1.52 (1.17)	1.21 (1.21)	3.02+ (1.62)	3.86* (1.65)	3.31* (1.65)	-3.77+ (1.99)	-2.73 (2.11)	-3.28 (2.19)	1.22 (1.70)	1.05 (1.72)	0.72 (1.73)	5.67* (2.83)	2.34 (2.36)	3.11 (2.40)
PPVT														
2003	5.61** (1.84)	5.77** (1.92)	10.70** (2.41)	11.09** (2.37)	11.21** (2.46)	0.46 (3.69)	1.27 (3.54)	1.74 (4.13)	2.00 (2.78)	2.13 (2.70)	2.73 (2.79)	11.55* (4.78)	9.98* (4.44)	8.33+ (4.90)
2004	3.03 (2.11)	1.92 (2.39)	6.93* (3.09)	7.54* (3.17)	7.71* (3.06)	-1.67 (4.50)	-2.82 (3.99)	-0.72 (4.24)	-2.35 (3.10)	-0.36 (3.10)	-2.03 (3.11)	5.30 (5.21)	4.77 (5.06)	4.07 (4.85)
2005	4.23* (1.73)	3.96* (2.00)	7.35** (2.69)	7.74** (2.77)	7.93** (2.84)	-1.46 (4.57)	-1.93 (4.12)	0.15 (4.06)	1.90 (2.56)	2.47 (2.46)	1.38 (2.61)	6.14 (4.27)	6.78 (4.40)	7.04 (4.54)
WJ-III Word														
2003	6.57** (1.67)	7.86** (1.85)	12.36** (2.46)	12.96** (2.37)	12.74** (2.38)	19.88** (3.80)	20.02** (3.42)	20.15** (3.74)	-0.33 (2.22)	0.03 (2.20)	-0.63 (2.26)	3.27 (3.29)	3.57 (3.05)	4.59 (3.21)
2004	0.19 (2.18)	-0.80 (2.50)	1.55 (3.45)	0.78 (3.45)	2.64 (3.38)	-1.41 (6.47)	-1.47 (6.10)	-0.69 (5.80)	-3.01 (3.11)	-1.78 (3.08)	-1.55 (3.14)	1.80 (4.93)	1.97 (4.58)	1.28 (4.12)
2005	1.48 (2.32)	1.73 (2.63)	5.55 (3.57)	6.15+ (3.53)	8.49* (4.09)	-3.30 (5.90)	-2.63 (5.44)	-4.01 (5.37)	-0.62 (3.31)	-0.39 (3.25)	-0.73 (3.44)	-6.13 (4.85)	-5.82 (4.69)	-6.00 (4.53)
WJ-III Applied														
2003	3.69* (1.55)	4.65** (1.75)	9.91** (2.62)	10.29** (2.59)	9.32** (2.56)	0.80 (2.99)	1.10 (2.81)	0.90 (2.93)	-0.32 (2.14)	0.38 (2.18)	0.31 (2.17)	1.51 (3.25)	0.59 (3.05)	1.93 (3.24)
2004	0.47 (1.47)	1.80 (1.80)	4.60 (2.80)	4.80+ (2.85)	5.14* (2.53)	3.04 (3.04)	3.31 (2.77)	2.55 (2.98)	-2.42 (1.81)	-1.26 (1.83)	-1.77 (1.90)	-1.85 (2.41)	-2.87 (2.52)	-1.79 (2.55)
2005	1.30 (1.27)	2.07 (1.46)	4.73* (2.07)	4.82* (2.01)	5.95** (2.25)	-0.37 (2.45)	0.07 (2.23)	0.72 (2.59)	-0.35 (1.94)	0.21 (2.00)	0.15 (2.02)	-0.55 (2.10)	-0.42 (2.03)	0.32 (2.15)
Social-behavioral Outcomes														
3-year-old cohort														
2003	0.04 (0.10)	0.03 (0.11)	-0.01 (0.14)	-0.05 (0.13)	-0.04 (0.14)	0.18 (0.22)	0.20 (0.22)	0.18 (0.23)	-0.03 (0.17)	-0.02 (0.17)	-0.04 (0.18)	0.11 (0.18)	0.09 (0.19)	0.10 (0.18)

Assignment Status	HS in Treatment Group vs. Specific Care in Control Group												
	Parental			Relative/Non-relative			Other Center-based			HS in Control			
ITT	TOT	OLS	IPW	PSM	OLS	IPW	PSM	OLS	IPW	PSM	OLS	IPW	PSM
2004	0.15	0.18	0.23 <sup>+</sup>	0.24 <sup>+</sup>	0.24 <sup>+</sup>	0.22	0.27	0.38 <sup>+</sup>	0.08	0.10	0.04	0.10	-0.02
	(0.10)	(0.11)	(0.14)	(0.14)	(0.14)	(0.23)	(0.22)	(0.22)	(0.18)	(0.20)	(0.17)	(0.20)	(0.19)
	0.24 <sup>*</sup>	0.22 <sup>+</sup>	0.18	0.25 <sup>+</sup>	0.23 <sup>+</sup>	-0.19	-0.25	0.04	0.51 <sup>*</sup>	0.58 <sup>*</sup>	0.60 <sup>**</sup>	0.41 <sup>+</sup>	0.36
2005	0.06	0.11	0.15	0.16	0.16	0.21	0.16	0.19	0.00	0.01	0.06	0.05	-0.02
	(0.10)	(0.11)	(0.14)	(0.15)	(0.15)	(0.22)	(0.22)	(0.23)	(0.14)	(0.14)	(0.14)	(0.18)	(0.16)
	0.12	-0.18	-0.13	-0.16	-0.11	-0.38 <sup>+</sup>	-0.41 <sup>*</sup>	-0.40 <sup>*</sup>	-0.10	-0.06	-0.11	-0.01	0.00
2003	0.10	0.11	0.14	0.14	0.14	0.20	0.20	0.20	0.18	0.18	0.18	0.21	0.20
	(0.10)	(0.11)	(0.14)	(0.14)	(0.14)	(0.20)	(0.20)	(0.20)	(0.18)	(0.18)	(0.18)	(0.21)	(0.20)
	-0.10	-0.31 <sup>**</sup>	-0.27 <sup>*</sup>	-0.30 <sup>*</sup>	-0.27 <sup>*</sup>	-0.26	-0.33 <sup>+</sup>	-0.46 <sup>*</sup>	-0.39 <sup>*</sup>	-0.40 <sup>*</sup>	-0.43 <sup>*</sup>	0.04	0.03
2004	0.10	0.11	0.13	0.14	0.13	0.20	0.19	0.20	0.19	0.20	0.20	0.19	0.19
	(0.10)	(0.11)	(0.13)	(0.14)	(0.13)	(0.20)	(0.19)	(0.20)	(0.19)	(0.20)	(0.20)	(0.19)	(0.19)
	-0.11	-0.22 <sup>+</sup>	-0.32 <sup>*</sup>	-0.35 <sup>*</sup>	-0.35 <sup>*</sup>	-0.24	-0.30 <sup>+</sup>	-0.33 <sup>+</sup>	-0.05	-0.12	-0.07	-0.07	-0.10
2005	0.11	0.12	0.15	0.15	0.15	0.19	0.18	0.20	0.19	0.19	0.20	0.22	0.22
	(0.11)	(0.12)	(0.15)	(0.15)	(0.15)	(0.19)	(0.18)	(0.20)	(0.19)	(0.19)	(0.20)	(0.22)	(0.22)
	-0.05	-0.14	-0.24	-0.30 <sup>+</sup>	-0.28 <sup>+</sup>	-0.39	-0.39 <sup>+</sup>	-0.47 <sup>*</sup>	0.15	0.06	-0.06	0.00	-0.17
2006	0.11	0.12	0.15	0.16	0.15	0.25	0.23	0.23	0.17	0.17	0.18	0.20	0.20
	(0.11)	(0.12)	(0.15)	(0.16)	(0.15)	(0.25)	(0.23)	(0.23)	(0.17)	(0.17)	(0.18)	(0.20)	(0.21)
	-0.35 <sup>**</sup>	-0.38 <sup>**</sup>	-0.50 <sup>**</sup>	-0.51 <sup>**</sup>	-0.55 <sup>**</sup>	-0.38 <sup>*</sup>	-0.35 <sup>+</sup>	-0.41 <sup>*</sup>	-0.22	-0.20	-0.28 <sup>+</sup>	-0.34 <sup>+</sup>	-0.29
2003	0.09	0.10	0.12	0.13	0.13	0.18	0.19	0.18	0.17	0.17	0.17	0.18	0.18
	(0.09)	(0.10)	(0.12)	(0.13)	(0.13)	(0.18)	(0.19)	(0.18)	(0.17)	(0.17)	(0.17)	(0.18)	(0.18)
	-0.13	-0.19 <sup>*</sup>	-0.23 <sup>*</sup>	-0.26 <sup>*</sup>	-0.24 <sup>*</sup>	-0.12	-0.16	-0.17	-0.20	-0.21	-0.25 <sup>+</sup>	-0.02	-0.03
2004	0.08	0.09	0.11	0.12	0.11	0.17	0.17	0.17	0.15	0.16	0.15	0.16	0.15
	(0.08)	(0.09)	(0.11)	(0.12)	(0.11)	(0.17)	(0.17)	(0.17)	(0.15)	(0.16)	(0.15)	(0.16)	(0.16)
	-0.21 <sup>*</sup>	-0.24 <sup>*</sup>	-0.20 <sup>+</sup>	-0.25 <sup>*</sup>	-0.24 <sup>*</sup>	-0.23	-0.26 <sup>+</sup>	-0.41 <sup>**</sup>	-0.26 <sup>+</sup>	-0.33 <sup>*</sup>	-0.29 <sup>*</sup>	-0.18	-0.17
2005	0.09	0.10	0.12	0.12	0.12	0.15	0.14	0.14	0.16	0.17	0.14	0.18	0.18
	(0.09)	(0.10)	(0.12)	(0.12)	(0.12)	(0.15)	(0.14)	(0.14)	(0.16)	(0.17)	(0.14)	(0.18)	(0.18)
	-0.14	-0.17 <sup>+</sup>	-0.17	-0.22	-0.25 <sup>+</sup>	-0.37 <sup>*</sup>	-0.35 <sup>*</sup>	-0.39 <sup>*</sup>	-0.06	-0.11	-0.14	0.01	-0.01
2006	0.09	0.10	0.13	0.14	0.14	0.19	0.17	0.18	0.16	0.16	0.16	0.18	0.18
	(0.09)	(0.10)	(0.13)	(0.14)	(0.14)	(0.19)	(0.17)	(0.18)	(0.16)	(0.16)	(0.16)	(0.18)	(0.18)
	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
4-year-old cohort	0.04	0.00	0.06	0.07	0.04	-0.34 <sup>+</sup>	-0.24	-0.25	0.04	0.09	0.04	-0.04	0.04
	(0.11)	(0.13)	(0.17)	(0.16)	(0.18)	(0.19)	(0.18)	(0.19)	(0.15)	(0.15)	(0.16)	(0.25)	(0.24)
	0.08	0.05	-0.09	-0.12	-0.08	0.25	0.19	0.19	0.16	0.18	0.19	-0.25	-0.22
Social Skills & Approaches to Learning	0.08	0.05	-0.09	-0.12	-0.08	0.25	0.19	0.19	0.16	0.18	0.19	-0.25	-0.22
	(0.11)	(0.13)	(0.17)	(0.16)	(0.18)	(0.19)	(0.18)	(0.19)	(0.15)	(0.15)	(0.16)	(0.25)	(0.24)
	0.08	0.05	-0.09	-0.12	-0.08	0.25	0.19	0.19	0.16	0.18	0.19	-0.25	-0.22

Assignment Status	HS in Treatment Group vs. Specific Care in Control Group															
	Parental			Relative/Non-relative			Other Center-based			HS in Control						
ITT	TOT	OLS	IPW	PSM	OLS	IPW	PSM	OLS	IPW	PSM	OLS	IPW	PSM	OLS	IPW	PSM
2005	0.05	-0.06	-0.12	-0.15	-0.05	-0.24	-0.25	-0.23	0.04	0.05	0.06	-0.02	0.00	0.09	0.09	0.09
	(0.10)	(0.12)	(0.14)	(0.14)	(0.14)	(0.24)	(0.20)	(0.19)	(0.17)	(0.17)	(0.18)	(0.22)	(0.21)	(0.23)		
2003	-0.16	-0.27*	-0.21	-0.27	-0.23	-0.72**	-0.72**	-0.74**	-0.23	-0.25+	-0.24+	0.21	0.25	0.09	0.09	0.09
	(0.11)	(0.11)	(0.16)	(0.17)	(0.16)	(0.20)	(0.18)	(0.18)	(0.14)	(0.14)	(0.14)	(0.25)	(0.24)	(0.25)		
2004	-0.06	-0.02	0.02	0.00	-0.02	-0.50+	-0.53*	-0.52*	0.06	0.04	0.08	0.05	0.11	0.18	0.18	0.18
	(0.11)	(0.12)	(0.16)	(0.16)	(0.16)	(0.30)	(0.27)	(0.26)	(0.16)	(0.16)	(0.16)	(0.27)	(0.26)	(0.25)		
2005	-0.08	-0.15	-0.15	-0.11	-0.18	-0.63*	-0.65**	-0.68**	-0.02	-0.05	-0.10	0.63*	0.50+	0.44	0.44	0.44
	(0.12)	(0.13)	(0.17)	(0.17)	(0.17)	(0.26)	(0.25)	(0.24)	(0.18)	(0.18)	(0.19)	(0.27)	(0.26)	(0.28)		
2003	-0.07	-0.19+	-0.27+	-0.28*	-0.28*	-0.08	-0.06	-0.11	-0.14	-0.15	-0.18	0.38*	0.40*	0.28+	0.28+	0.28+
	(0.09)	(0.10)	(0.14)	(0.13)	(0.13)	(0.18)	(0.16)	(0.16)	(0.14)	(0.14)	(0.14)	(0.17)	(0.16)	(0.16)		
2004	0.12	0.02	-0.08	-0.06	-0.11	0.00	-0.06	-0.03	0.12	0.10	0.09	0.21	0.24	0.14	0.14	0.14
	(0.10)	(0.11)	(0.15)	(0.15)	(0.15)	(0.25)	(0.23)	(0.21)	(0.15)	(0.15)	(0.15)	(0.19)	(0.19)	(0.19)		
2005	0.03	-0.02	0.00	0.06	-0.04	-0.41	-0.49+	-0.61*	0.06	0.06	0.01	0.32	0.20	0.20	0.20	0.20
	(0.10)	(0.11)	(0.16)	(0.15)	(0.16)	(0.26)	(0.26)	(0.25)	(0.15)	(0.15)	(0.15)	(0.20)	(0.19)	(0.20)		

Notes: coefficients of Head Start with standard errors in parentheses reported in table;

\*\* p<0.01,

\* p<0.05,

+ p<0.10.