



Published in final edited form as:

Dev Psychol. 2011 January ; 47(1): 134–152. doi:10.1037/a0020784.

Head Start and Urban Children's School Readiness: A Birth Cohort Study in 18 Cities

Fuhua Zhai,

School of Social Welfare, Stony Brook University

Jeanne Brooks-Gunn, and

Pediatrics Department, Teachers College and the College of Physicians and Surgeons, Columbia University

Jane Waldfogel

School of Social Work, Columbia University

Abstract

We used longitudinal data from a birth cohort study, the Fragile Families and Child Wellbeing Study, to investigate the links between Head Start and school readiness in a large and diverse sample of urban children at age 5 ($N = 2,803$; 18 cities). We found that Head Start attendance was associated with enhanced cognitive ability and social competence and reduced attention problems but not reduced internalizing or externalizing behavior problems. These findings were robust to model specifications (including models with city-fixed effects and propensity-scoring matching). Furthermore, the effects of Head Start varied by the reference group. Head Start was associated with improved cognitive development when compared with parental care or other nonparental care, as well as improved social competence (compared with parental care) and reduced attention problems (compared with other nonparental care). In contrast, compared with attendance at pre-kindergarten or other center-based care, Head Start attendance was not associated with cognitive gains but with improved social competence and reduced attention and externalizing behavior problems (compared with attendance at other center-based care). These associations were not moderated by child gender or race/ethnicity.

Keywords

Head Start; school readiness; urban children; propensity score matching

Since its inception in 1965, Head Start has been the single largest publicly financed early childhood education and care program in the United States. Head Start's primary goal is to improve the school readiness of children from low-income families by delivering high-quality and comprehensive early education services to preschool age children, in particular, 3- and 4-year-olds. However, throughout its history, the success of the program in meeting this goal has been debated (Styfco & Zigler, 2004) and continues to be (Besharov & Call, 2009; Nisbett, 2009), even though recent well-designed observational studies (e.g., Currie & Thomas, 1995, 1999; Garcés, Thomas, & Currie, 2002; Ludwig & Miller, 2007) have reported significant short- and long-term benefits of Head Start. Most recently, the only randomized experiment to date, the Head Start Impact Study, reported short-term benefits

(U.S. Department of Health and Human Services, Administration for Children and Families [ACF], 2005), although these were not maintained in the longer term (ACF, 2010).

One challenge that is common to all nonexperimental Head Start studies is to account adequately for selection bias, given that the program by design serves children who are poor. Although recent studies have had stronger research designs, including the one random assignment experiment, to address selection bias, a further challenge is that many studies have not clearly defined the reference group to which Head Start children were being compared, even though children not attending Head Start might attend a variety of alternative care settings (V. E. Lee, Brooks-Gunn, & Schnur, 1988; ACF, 2005, 2010). The lack of a clear reference group may contribute to the variation in findings across studies given the considerable variations in child care programs and policies across localities and time periods (Ludwig & Phillips, 2007, 2008; Rigby, Ryan, & Brooks-Gunn, 2007; Waldfogel, 2006).

In this study, we used data from a large longitudinal birth cohort study of primarily low-income children in urban areas, the Fragile Families and Child Wellbeing Study (FFCWS; 2008a), to investigate the effects of Head Start participation on children's school readiness. The fact that our sample was mainly made up of disadvantaged families helped address some of the issues with regard to selection bias, but to further address possible selection bias, we adopted several different analytic approaches, including ordinary least squares (OLS) regressions with a rich set of controls, city-fixed effects approaches, and propensity score matching models. In addition, we were able to control for children's earlier developmental outcomes (i.e., at age 3 when almost none of the children had attended Head Start), which has not been possible in most of the previous research. In common with prior studies, we first examined the effects of Head Start by comparing Head Start participants to all nonparticipants (regardless of what their child care arrangements were). Then, to address the problem of the lack of clarity with regard to the reference group, we compared children who attended Head Start with children who attended specific types of other care arrangements separately, including parental care, prekindergarten, other center-based care, and other nonparental care. We also analyzed whether the effects of Head Start were moderated by child gender and race/ethnicity.

Background and Prior Research

The aims of Head Start are to promote multiple aspects of children's school readiness (e.g., cognitive development, learning skills, social competence, health, and nutrition) through the provision of comprehensive and high-quality services, including early education and development; parental involvement; and medical, dental, mental health, and nutritional programs as well as other social services (Aughinbaugh, 2001; Blau, 2001; Smolensky & Gootman, 2003; ACF, 2009). Descriptive statistics based on data collected in the spring of the programs in the Head Start Impact Study suggest that compared with teachers in other center-based classrooms serving low-income families, Head Start teachers tend to be less harsh, less detached, less permissive, more sensitive, and more likely to encourage children to be independent and to promote children's active involvement in learning and cooperative behaviors with teachers and peers (ACF, 2005). Compared with parents of nonparticipants, Head Start parents are more emotionally supportive, more likely to read to their children, less detached, and less likely to use physical discipline and have overall better quality of home environment (ACF, 2005).

A large body of observational and experimental studies has consistently shown that high-quality care, especially that measured by process indicators, such as caregivers' warmth, sensitivity, responsiveness, consistency, and stimulation of interactions, is linked to

children's cognitive development, social competence, and attention, especially for children from disadvantaged families (although the size of these effects is debated; see K. Lee, 2005; Magnuson & Waldfogel, 2005; National Institute of Child Health and Human Development [NICHD] Early Child Care Research Network & Duncan, 2003; Votruba-Drzal, Coley, & Chase-Lansdale, 2004; Waldfogel, 2006). Thus with better school and home environments and more competent teachers and parents, Head Start participants are expected to have better opportunities and resources to prepare for school and achieve better cognitive and social-emotional development than their nonparticipant peers.

Nevertheless, findings on the effects of Head Start have varied considerably across studies, ranging from negative or no effects to substantial positive long-term effects of Head Start (see reviews by Aughinbaugh, 2001; Blau, 2001; Currie, 2001; Smolensky & Gootman, 2003). One of the main reasons contributing to the divergent findings is the issue of selection bias in observational studies of Head Start (given that in 40 years, there has been only one randomized evaluation). As discussed earlier, the target population of Head Start is disadvantaged preschool-age children, who overall tend to have worse developmental outcomes than their advantaged peers even before attending Head Start (Currie, 2005; V. E. Lee et al., 1988; Reid, Webster-Stratton, & Baydar, 2004). In addition, at the time they enter Head Start, most children have early literacy and math skill scores that are well below national averages (ACF, 2006). Thus, it is likely that many observable and unobservable factors relevant to disadvantaged families (e.g., low income, single parenthood, low parental education, and high parental stress) could affect both selection into Head Start program and outcomes of participants. As a result, simply comparing the outcomes of Head Start participants with those of nonparticipants could bias the estimation of the "true" effects of Head Start.

The most recent observational studies have been designed to address the issue of selection bias in Head Start research (e.g., use of family fixed effects, regression discontinuity, and propensity score matching models) and have documented some substantial short- and long-term benefits of Head Start for participants. Some examples include improvements in cognitive development, school achievement, social skills, college attendance, medical care, and health status, as well as reductions in grade retention, special education, high-school dropout rates, teen pregnancies, delinquency, and criminal activities (see reviews and research by Blau, 2001; Currie, 2001; Currie & Thomas, 1995, 1999; Deming, 2009; Garces et al., 2002). For instance, using sibling comparison or family fixed effects and data from the National Longitudinal Survey of Youth, Currie and Thomas (1995) found that Head Start participation was associated with positive gains in test scores of children 8–10 years old compared with their siblings who either had not attended preschool or had attended other preschools. Using the same method and data, Currie and Thomas (1999) found that, on average, Head Start closed at least one quarter of the gap in test scores and two thirds of the gap in the probability of grade retention between Hispanic children and non-Hispanic White children, while Deming (2009) found a long-term impact for 0.23 standard deviations of Head Start enrollment on a summary index of young adult outcomes. Ludwig and Miller (2007) adopted a regression discontinuity method and found a large drop in mortality rates and increases in high school completion and college attendance among Head Start participants.

A recent experimental study, the Head Start Impact Study, was the first in which the impacts of Head Start were assessed with a random assignment method. Children whose families applied to Head Start programs that were oversubscribed were randomly assigned to either receive Head Start (i.e., the treatment group) or be placed on a waiting list (i.e., the control group). After 1 year of participation, the 3- and 4-year-old children who were randomly assigned to Head Start programs had significantly higher scores in reading, writing,

vocabulary, and parent-reported literacy skills than children in the control group (ACF, 2005). However, the magnitude of the effects was generally modest, particularly when compared with effects that had been documented for other smaller scale but high-quality model early interventions (e.g., Perry Preschool, Abecedarian, and the Infant Health and Development Program; with short-term effect sizes on cognitive outcomes ranging from 0.35 to 0.97) or more recent evaluations of universal state prekindergarten programs (e.g., those in Oklahoma and West Virginia; with effect sizes on academic outcomes ranging from 0.26 to 0.80)¹ (Barnett & Hustedt, 2005; Barnett, Lamy, & Jung, 2005; Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001; Currie, 2001; Gormley, Gayer, Phillips, & Dawson, 2005; Karoly, Kilburn, & Cannon, 2005; Kilburn & Karoly, 2008; Ludwig & Phillips, 2007, 2008; Schweinhart et al., 2005).

In addition to selection bias, a further challenge in Head Start research is the change of counterfactual across studies, which may have been an important factor contributing to some variations in findings of Head Start effects as well as smaller effects compared with the older model early interventions. Since the availability and quality of child care programs vary considerably across time periods and localities, the composition of the non-Head Start reference group has likely varied as well. For example, few 3- and 4-year-olds were in any form of preschool in the 1960s or 1970s when the Head Start programs as well as the model early interventions evaluated in many studies were provided, meaning that many children in the reference group would have received no preschool education. In contrast, among contemporary cohorts, more than 40% of 3-year-olds and nearly 70% of 4-year-olds are now in some form of school or center-based care or education on a regular basis (U.S. Census Bureau, 2008; Waldfogel, 2006). The quality of Head Start and its alternatives may have been changing over time as well (Ludwig & Phillips, 2007, 2008). In addition, child care programs and policies (e.g., access, funding levels, program standards, and teacher quality) also vary considerably across localities and states (Rigby et al., 2007). Therefore, depending on the specific time periods and localities of sampling, the care arrangements children in the non-Head Start control group had access to and received might vary substantially across studies. Research has shown that the type and quality of care arrangements are closely related to children's developmental outcomes.² Thus, it is critical to clearly define the reference groups for Head Start; otherwise, the estimated effects of Head Start could vary substantially across studies depending on the sampling of children who had other care arrangements.

Unfortunately, with a few exceptions (e.g., Currie & Thomas, 1995, 1999; Garces et al., 2002; V. E. Lee et al., 1988), the estimates of Head Start effects in many prior studies³ were obtained through comparison of children participating in Head Start with all other children, which meant that the reference group included children experiencing a mixture of alternative care settings ranging from exclusively parental care to other center-based care (V. E. Lee et

¹It should be noted that the effects of Head Start reported in the Head Start Impact Study (and many model early interventions) were from intention-to-treat estimates, while those from state universal pre-K programs were from treatment-on-treated (TOT) estimates (see the detailed discussions in Ludwig & Phillips (2007, 2008) and ACF, 2005).

²For example, children who attended center-based care tended to have higher cognitive and social skills but more behavior problems than children who received parental or relative care (see reviews and research by Baydar & Brooks-Gunn, 1991; Clarke-Stewart & Allhusen, 2005; Côté, Borge, Geoffroy, Rutter, & Tremblay, 2008; Hill et al., 2002; Magnuson & Waldfogel, 2005; NICHD Early Child Care Research Network, 2003; NICHD Early Child Care Research Network, 2005). It is also well documented that there are sizeable cognitive benefits of attending prekindergarten, especially for disadvantaged children (Gormley, 2008; Gormley, Phillips, & Gayer, 2008; Magnuson et al., 2007). In contrast, generally little or no benefit has been found for children who attended other types of nonparental care, and children who were in exclusive parental care right before kindergarten have been found to lag behind their peers, particularly in terms of cognitive development (see reviews in Smolensky & Gootman, 2003; Waldfogel, 2006).

³It should be noted that the Head Start Impact Study created the benchmark since its reference group was clearly defined as children who were just like children in the treatment group except that by chance they had not been admitted into the program. However, since the randomization was conducted over the eligibility to enroll in Head Start programs but not specific other care arrangements, the direct comparison of Head Start to the reference groups of other care arrangements would be internally invalid.

al., 1988; ACF, 2005). The lack of an explicitly defined reference group in most prior Head Start research makes it difficult to draw firm conclusions from the past generation of observational studies (as well as the contemporary experimental study).

The Present Study

To address these challenges and unresolved questions in prior research, we used data from the FFCWS to examine the effects of Head Start participation on the school readiness of a large and diverse sample of children who were born to low-income families in large U.S. cities in the late 1990s. Multiple dimensions of children's school readiness at age 5 were analyzed, including cognitive development, social competence, and attention and behavior problems. To address the issue of selection bias, we adopted several different analytic approaches, including OLS regressions with a rich set of pretreatment controls (i.e., child demographics and earlier ability or behavior), OLS regressions with city-fixed effects, and propensity score matching models. Like previous researchers, we first focused on the effects of Head Start compared with any other care arrangements. Then, to further explore the effects of Head Start compared with specifically defined reference groups, we conducted separate analyses in which children who attended Head Start were compared with children experiencing specific other types of care arrangements, including parental care, prekindergarten, other center-based care, and other nonparental care. Finally, given prior mixed findings as to whether the effects of Head Start might be moderated by child gender or race/ethnicity, we also carried out supplemental analyses to examine both gender and race/ethnicity.

Method

Data and Sample for Analysis

We used data on a birth cohort of new parents and children from 18 large U.S. cities from the FFCWS (2008b; Reichman, Teitler, Garfinkel, & McLanahan, 2001). As a national longitudinal study of a large and diverse sample of predominantly low-income urban children, the FFCWS researchers used a stratified random sample of all U.S. cities with 200,000 or more people to randomly select participating cities on the basis of three policy and labor market indicators (i.e., welfare generosity, strength of child support system, and strength of labor market). If the selected cities had five or fewer birthing hospitals, all the hospitals were then selected; the hospitals in other cities were selected starting from the largest ones until the sample size was met or selected randomly if there were dozens of large birthing hospitals (see Reichman et al. 2001, for details).⁴

Overall, the FFCWS sample contained 4,242 children born in these cities between 1998 and 2000. Baseline interviews were conducted in person at the hospitals shortly after the focal child was born, followed by telephone interviews when the focal child was approximately 1, 3, and 5 years old. The measures of school readiness in the present study were extracted from a collaborative study of FFCWS, the In-Home Longitudinal Study of Pre-School Aged Children, which included in-depth interviews of primary caregivers (typically the child's mother) and in-home direct assessments when the focal child was 3 and 5 years old, focusing on parenting, child health, and development (FFCWS, 2008b).

Of the original 4,242 children, 647 children (15%) were not followed up in the 5-year phone interview, while 743 children (18%) did not participate in the 5-year in-home study. An

⁴Similar to most other studies using the FFCWS data, in the multivariate analyses, we did not use sampling weights; we did control for all the variables that were used for weights, including mother's marital status, education, age, and race/ethnicity (see Reichman et al., 2001, and FFCWS, 2008a or more information regarding the use of sampling weights in analyses with FFCWS data).

additional 49 children (1%) were dropped from the analysis because they had missing information on care arrangements or on outcome measures. These restrictions narrowed the sample for analysis to 2,803 children (i.e., 66% of the original sample).

In spite of these restrictions, our analysis sample was similar to the original FFCWS sample on most demographics and family background variables. Significant differences between children in our analysis sample and children who were not included overall tended to be modest.⁵

However, our analysis sample still consisted primarily of children in low-income families and was very diverse in terms of race/ethnicity. As shown in Table 1, nearly half (49%) were non-Hispanic Black children, followed by Hispanics (20%), non-Hispanic Whites (17%), and children of biracial or other racial/ethnic groups (14%). Approximately 36% of children lived in poverty when they were born. In addition, 71% of children had mothers who were not married when children were 3 years old, and more than one quarter (27%) of children had mothers with less than high school education.

Since the Head Start Impact Study included nationally representative Head Start children, we also conducted a comparison of Head Start participants in both studies. As shown in Appendix Table B, the Head Start children in our analysis sample were also comparable to those in the Head Start Impact Study in terms of gender; likelihood that mother was a teen mom; mother's age, education, employment, and self-reported health status; and household income. Differences between the Head Start children in our sample and those in the Head Start Impact Study (e.g., child race/ethnicity and parents' marital status) were mainly because the sampling of FFCWS heavily focused on nonmarital births in large U.S. cities while the Head Start Impact Study was representative of Head Start children nationwide.

Measures

Outcome variables—Aspects of children's school readiness included cognitive development, social competence, and attention and behavior problems at age 5. Specifically, children's cognitive development was measured by the Peabody Picture Vocabulary Test (3rd ed.; PPVT-III; Dunn, Dunn, & Dunn, 1997) and the Woodcock-Johnson Psycho-Educational Battery-Revised (WJ-R) Letter-Word Identification. As a widely used receptive vocabulary test, the PPVT-III measures children's language and cognitive ability. Another of the most widely used instruments for assessing cognitive abilities and achievement in children, the WJ-R Letter-Word Identification, measures children's reading identification skills (Mather & Jaffe, 1996).

Children's social competence was measured by a subset of items asked of the mother from the Adaptive Social Behavior Inventory (ASBI) Express Subscale (Hogan, Scott, & Bauer, 1992). The FFCWS 5-year in-home direct assessments contained 12 items ($\alpha = .82$) of ASBI Express Subscale that measured children's positive behaviors. Examples of these items included that the child understood others' feelings, was open about what he or she wanted, would join a group of children playing, was confident with other people, and tended to be proud of things he or she did.

The attention and behavior problems were assessed by subscales from the Child Behavior Checklist for Ages 1.5-5 (CBCL/1.5-5; Achenbach & Rescorla, 2000). In our analyses,

⁵As shown in Appendix Table A, among Head Start participants, those in the analysis sample tended to be more disadvantaged (e.g., more likely to have mothers with lower cognitive ability scores and lower family income) than those who were excluded due to attrition, while the comparison of nonparticipants in the analysis sample with those out of the sample tended to be mixed in terms of demographics and family background. As a result, it was unclear whether our findings might underestimate or overestimate the effects of Head Start (if those are greater for children who are the most disadvantaged).

based on the CBCL/1.5–5 items reported by mothers in the FFCWS 5-year in-home assessments, we used three aggregated subscales to measure children’s attention and behavior problems. Specifically, the subscale of attention problems was aggregated from 11 items ($\alpha = .73$). Some examples of attention problems included that child could not concentrate for long, could not sit still, was nervous, was confused, day dreamed or got lost in thoughts, was impulsive, and stared blankly. The subscale of internalizing behavior problems was aggregated from 22 items ($\alpha = .76$).⁶ Examples of items included that the child cried a lot, was too fearful or anxious, was sad or depressed, felt worthless or inferior, complained of loneliness, felt too guilty, worried, and was withdrawn. Externalizing behavior problems were indexed by 30 items ($\alpha = .85$). Some examples were that child lied or cheated, ran away from home, set fires, stole at home or outside the home, argued a lot, was disobedient at home or at school, and got into many fights.

These measures of young children’s school readiness have been extensively used in large-scale observational studies and policy evaluations as well as in smaller efficacy trials of educational and clinical interventions (e.g., Berger, Brooks-Gunn, Paxson, & Waldfogel, 2008; Hill, Brooks-Gunn, & Waldfogel, 2003; Love et al., 2005; Markowitz et al., 2006; Meadows, McLanahan, & Brooks-Gunn, 2007; ACF, 2005; Yeung, Linver, & Brooks-Gunn, 2002). We presented the summary statistics of these measures at child’s ages 3 and 5 in Table 1. In the analyses, to make it easier to interpret the estimates and to compare the effect sizes of Head Start across outcome variables as well as to compare them with findings in other studies, we standardized all the measures of children’s school readiness to have a mean of 0 and a standard deviation of 1.

Child care arrangements right before kindergarten—In the 5-year follow-up interview of the FFCWS core study when children on average were about 5 years old, parents were asked about the focal child’s current child care arrangements or, if the child was already in kindergarten, about the child’s child care arrangements from January to May of that year. Specifically, parents were asked whether the focal child was currently attending center-based care on a regular basis if he or she was not in kindergarten or whether the child attended center-based care on a regular basis between the beginning of January through the end of May of the survey year if he or she was currently in kindergarten (i.e., kindergarten was not counted as a child care arrangement). If the answer was “yes,” parents were asked to choose what type of program that the child was attending, or attended, most regularly from a forced-choice list of options, including day care center, nursery school, preschool, Head Start program, and prekindergarten. We used parents’ responses to these choices as the child’s main center-based care arrangement right before kindergarten and, following prior research (Magnuson, Ruhm, & Waldfogel, 2007), recoded them as either Head Start,⁷ prekindergarten, or other center-based care (this latter category combined what parents referred to as day care centers, nursery schools, and preschool programs).

If the answer to the regular center care attendance question was “no,” parents were asked whether the child was cared by someone other than the custodial parents for at least 8 hr every week for a month or more (again asking about the current arrangement for children

⁶The internalizing behavior problem subscale also included two items that were included in the subscale of attention problems (i.e., that the child was nervous and that the child stared blankly).

⁷Overall only 48 children (i.e., 1.7% of the analysis sample) attended Head Start or Early Head Start (a program that serves children younger than 3 years old) in the FFCWS 3-year survey when children on average were about 3 years old. In the 5-year FFCWS survey, 12 of these children were in Head Start, 18 in prekindergarten, 14 in other center-based care, one in other nonparental care, and three in parental care. Since we focused on children’s Head Start attendance right before kindergarten, children who previously attended Head Start, including those who had been in the FFCWS 3-year survey but were not in Head Start right before kindergarten in the FFCWS 5-year survey, were coded as nonparticipants of Head Start. We also re-estimated the models excluding all children who were previous Head Start participants, and the results were almost identical to those reported.

not yet in kindergarten or about the arrangement in January through May of that year for children already in kindergarten). We used the answers to this question to define children who were not attending center-based care on a regular basis but were receiving care from others than their parents for at least 8 hr per week for a month or more as receiving other nonparental care, which included grandparent care, other relative care, nonrelative care, family care, and other care. The remaining children who were neither attending center-based care on a regular basis nor receiving other nonparental care for at least 8 hr per week for a month or more (again, at the time of the age-5 interview or, for those already in kindergarten, from January to May of that year) were categorized as receiving parental care (i.e., having exclusive parental care or nonparental care, if any, for less than 8 hr per week).

In the analysis,⁸ we first focused on the effects of Head Start participation compared with any other care arrangements (i.e., Head Start vs. non-Head Start). Following these analyses, we further compared Head Start to the other specific types of child care arrangements (i.e. prekindergarten, other center-based care, other nonparental care, or parental care).

Other covariates—The FFCWS followed children from birth, collecting rich data on child demographics and family background. In the analyses, we included an extensive set of covariates that were measured at the child's birth and at ages 1 and 3 and that might have affected both the child's probabilities of Head Start participation right before kindergarten and his or her school readiness outcomes at age 5. Specifically, child demographics included gender, age at assessment, and race/ethnicity and data on whether the child was the mother's first child, had a low birth weight, or had fair or poor health status reported by the mother. In the analyses of outcome measures at age 5, we also controlled for children's corresponding pretreatment scores of PPVT-III, social competence, and attention and behavior problems, which were collected in the 3-year in-home direct assessments.⁹

We also controlled for the mother's demographics and household characteristics in the analyses. The mother's demographics included age, whether her first child was born when she was 18 years old or younger, relationships with the child's father in the 3-year survey, employment at child's age 3, and education. We also controlled for two measures collected in the FFCWS 3-year core study, the mother's cognitive ability and depression, that were likely to be associated with children's cognitive development and behavior problems (Berger et al., 2008; Lanzi, Pascoe, Keltner, & Ramey, 1999). The mother's cognitive ability was measured by the Similarities subtest of the Wechsler Adult Intelligence Scale- Revised (WAIS-R), which measures adult intelligence through assessment of verbal concept information and reasoning abilities (Wechsler, 1981). The mother's depression was measured by whether the mother had felt sad, depressed, or anxious for 2 or more weeks in the past year.

In addition, we also included variables for the mother's parenting and household cognitively stimulating materials and household income in the analyses. Mother's parenting and household cognitively stimulating materials were drawn from subscales of the Home Observation for Measurement of the Environment (HOME), which has been widely used to measure the quality of home environment for learning and cognitive stimulation (Bradley, 1993). On the basis of previous research (e.g., Berger et al., 2008; Bradley & Corwyn, 2007; Fuligni, Han, & Brooks-Gunn, 2004; Leventhal, Martin, & Brooks-Gunn, 2004), we used

⁸In supplemental models, we also controlled for earlier child care experiences prior to age 5, including those at age 3 as well as those at age 1. The effects of Head Start participation right before kindergarten in those supplemental models were identical to those from models without these earlier child care experiences being controlled.

⁹WJ-R scale was not assessed at age 3, and thus the pretreatment scores of PPVT-III were used for the analysis of WJ-R. We refer to the age-3 measures as pretreatment, since most children who subsequently attended Head Start had not started attending the program at the time of the age-3 assessment.

three aggregated subscales from data collected in the 3-year in-home assessments: harsh parenting, maternal responsiveness, and cognitively stimulating materials available to child.¹⁰ Finally, we controlled for household income relative to poverty threshold at the time of the child's birth and age 3.¹¹

Analytic Strategies

In the analysis, we first used OLS regressions with a series of increasingly controlled models. We started with a model that included Head Start participation status (1 = yes, 0 = no) and child demographics. We then added child pretreatment scores on the relevant outcome variables at age 3 in the models. We further added mother and family covariates into the models to see whether controlling for those potential selection factors changed the results. The descriptive analyses within cities showed that the distribution of child care arrangements was different across cities (table available from the authors upon request). Thus, as a robustness check, we further controlled for city-fixed effects to account for the heterogeneity across cities in Head Start programs, the availability and usage of other types of child care arrangements, and other contextual factors that could affect children's participation in Head Start as well as their developmental outcomes. The model with city-fixed effects is specified in Equation 1:

$$O_{imc} = \beta_0 + \beta_1 HS_{imc} + \beta_2 X_{imc} + \beta_3 M_{ic} + \varphi_c + \xi \quad (1)$$

where O_{imc} represents a school readiness outcome for individual child i with mother m in city c ; HS_{imc} indicates whether the child attended Head Start ($HS_{imc} = 1$) or not ($HS_{imc} = 0$) in the year right before kindergarten; X_{imc} is a vector of child characteristics and pretreatment scores; M_{ic} stands for a vector of mother and family characteristics; φ_c denotes city fixed effects; and ξ is a random error term. Huber–White robust standard errors clustered at city level were adopted.

To further address the issue of selection bias, we employed a propensity score matching method. Propensity score matching has been increasingly used in recent years to evaluate the effects of early childhood intervention programs (see research and reviews by Berger et al., 2008; Hill et al., 2003; Hill, Waldfogel, & Brooks-Gunn, 2002; Hill, Waldfogel, Brooks-Gunn, & Han, 2005; Schneider, Carnoy, Kilpatrick, Schmidt, & Shavelson, 2007; Shonkoff & Philips, 2000). In a conventional propensity score matching approach, observed pretreatment covariates are used to estimate the probability of being in the treatment group (i.e., the propensity score). Then, for each member in the treatment group, one or more “matched” members in the control group are identified with the closest propensity scores by various ways of matching such as nearest neighbor, radius, kernel, and Mahalanobis methods (Dehejia & Wahba, 1999, 2002; Gibson, 2003; Hill et al., 2002; Rosenbaum &

¹⁰The harsh parenting subscale consisted of five items indicating whether mother shouted at child, expressed hostility toward child, slapped or spanked child, scolded child, or restricted child more than three times. The maternal responsiveness measure included six items indicating whether mother spontaneously vocalized to or praised child at least twice, responded verbally to child's vocalizations, told child the name of an object or person, conveyed positive feelings toward child, or caressed child at least once. The subscale of cognitively stimulating materials was aggregated from 10 items, which included questions about the toys and books available to child.

¹¹A small proportion of children in the analysis sample had missing data on pretreatment scores or other covariates. We adopted two different approaches to address the missing data. First, for children who were not assessed at age 3 but were assessed at age 5 ($n = 293$; 10% of the full sample), we employed a multiple imputation approach, using “uvis” (univariate imputation sampling) in Stata 10 that implements multiple imputation by chained equations and a bootstrap method that estimates regression coefficients in a bootstrap sample of the nonmissing observations (Hill et al., 2002; Royston, 2005; Rubin, 1987; van Buuren, Boshuizen, & Knook, 1999). Second, to fill in the missing data on child demographic and family background covariates (ranging from 1% to 8%), we created a category of “missing” for categorical variables to flag those observations with missing values; for continuous variables, we replaced the missing values with the means of the nonmissing observations and created a dummy variable to indicate whether the values of observations were imputed by the means. In both cases, the categories that indicate missing observations were always included in the regression models.

Rubin, 1985). These matched children with similar propensity scores can be conceptualized as being analogous to children randomly assigned to the treatment or control groups in an experiment (Hill et al., 2002; Schneider et al., 2007), although the success of propensity score methods in approximating experimental results has been debated (see Dehejia, 2005; Smith & Todd, 2005; and also review by Rutter, 2007). It should be noted that propensity score matching methods require the assumption of selection only on observables but not on unobservables (Hill et al., 2005).

In this study, we adopted a three-stage propensity score matching method to identify children who did not attend Head Start but were comparable to Head Start participants in terms of their demographics and family background at age 3. In the first stage, we used the child, mother, and family covariates that were collected when the focal child was 3 years old, as detailed earlier, to predict the each child's probability of attending Head Start right before kindergarten (i.e., the propensity score). The propensity of Head Start attendance (HS) for child i with mother m in city c was estimated by a logit model specified in Equation 2:

$$\text{logit}(HS_{imc}) = \beta_0 + \beta_{imc} X_{imc} + \varphi_c \quad (2)$$

where X_{imc} denotes a sum of child, mother, and family characteristics and φ_c represents city-specific fixed effects. The inclusion of city-fixed effects in the predictive model is important to capture both observed and unobserved city level factors that might affect Head Start participation, such as the eligibility, funding, and implementation of Head Start programs, as well as other contextual variables, such as the availability and usage of other types of child care arrangements.

In the second stage, we matched each Head Start participant with a child who did not attend Head Start but lived in the same city and had the closest propensity score, using a one-to-one nearest neighbor matching method with replacement. Matching with replacement allows each treatment unit to be matched with the nearest control unit (i.e., some control units could be used more than once) and thus minimizes bias (Dehejia & Wahba, 1999, 2002; Gibson, 2003). In addition, we used a "common support" option to limit Head Start participants to those whose propensity scores were not higher than the maximum or less than the minimum propensity scores of nonparticipants (Leuven & Sianesi, 2003). Under the assumption that the predictors in Equation 2 were the only confounding variables, each pair of matched children who had similar propensity scores thus was comparable to other pairs in terms of the likelihood of Head Start attendance.

In the third stage, the effects of Head Start were estimated by the regression-adjusted differences in children's outcomes between Head Start participants and nonparticipants within the same pairs matched in the second stage. Regression-adjusted differences reduce variance by the inclusion of covariates in the models and thus can reduce the uncertainty in outcomes and increase the chance of detecting significant treatment effects (Gibson, 2003; Hill et al., 2002, 2003; ACF, 2005). To estimate the effects of Head Start, we used OLS regressions with pair-fixed effects, controlling for the pretreatment child, mother, and family covariates, as presented in Equation 3:

$$O_{imcp} = \beta_0 + \beta_1 HS_{imcp} + \beta_2 X_{imcp} + \beta_3 M_{icp} + \psi_p + \varepsilon \quad (3)$$

which includes the school readiness measure (O_{imcp}) of child i with mother m in city c within matched pair p ; his or her Head Start participation status (HS_{imcp}), demographics and pretreatment scores (X_{imcp}), and characteristics of mother and family (M_{icp}); the matched

pair fixed effects (ψp); and a random error term (ξ). To get correct inference, we clustered standard errors at the level of matched pairs. We did not include city-fixed effects in these models, but such effects were controlled for since pairs of children were matched within cities.

In the analysis, we first examined the average effects of Head Start compared with any non-Head Start care arrangements (i.e., Head Start vs. non-Head Start), using the models previously specified. To further understand the effects of Head Start compared with other specific types of child care arrangements (rather than just the average effects of Head Start in the first step), we applied the analytic models separately to subsamples consisting of children who had attended Head Start and one of the specific reference groups of child care arrangements, including parental care, prekindergarten, other centerbased care, and other non-parental care.¹² Children who experienced other care arrangements might have had different probabilities to participate in Head Start programs due to their different demographics and family backgrounds. Therefore, in the propensity score matching approach, limiting the analysis to a sample with Head Start and another specific reference group of child care arrangement could identify children who received this type of care and were comparable to Head Start participants. This analytic strategy further accounted for the issue of selection bias and thus could provide more sound estimates of the effects of Head Start compared with other specific types of child care arrangements. Finally, to investigate whether the effects of Head Start were moderated by child gender and race/ethnicity, we added interaction terms between Head Start participation and child gender and race/ethnicity, respectively, to the models.

Results

Descriptive Statistics and Results of Propensity Score Matching

Table 1 presents the descriptive statistics of pretreatment covariates as well as the unstandardized outcome measures at age 5 by children's Head Start participation and matching status. The second column shows results for the full sample in analysis ($N = 2,803$), followed by descriptive statistics among non-Head Start children before propensity score matching ($n = 2,417$), Head Start participants ($n = 386$) and non-Head Start children after propensity score matching ($n = 375$).¹³ Two-tailed t statistics were used to test the mean differences between Head Start participants and nonparticipants before and after propensity score matching.

In terms of child care arrangements right before kindergarten, as presented in Table 1, overall 14% of children in the full analysis sample attended Head Start programs. Other children received parental (16%), prekindergarten (25%), other center-based (37%), or other nonparental (8%) care. Among children who did not attend Head Start programs, compared with children before propensity score matching, children after matching tended to receive parental care (24% vs. 19%) or other nonparental care (11% vs. 9%) rather than prekindergarten (27% vs. 29%) or other centerbased (38% vs. 43% care).

Table 1 also shows that Head Start participants and nonparticipants before propensity score matching were significantly different on most pretreatment variables. Consistent findings from prior studies (e.g., V. E. Lee et al., 1988), non-Head Start children before matching tended to be more advantaged on most child and family covariates compared with Head

¹²Common support and matching with replacement were also adopted in the propensity score matching analyses for each of the subsamples.

¹³Overall 11 Head Start participants (i.e., 2.8% of 386 participants) were not in the area of common support and thus were not matched to nonparticipants.

Start participants. For example, nonparticipant children were more likely to be non-Hispanic White and biracial/other rather than non-Hispanic Blacks, to be the first children of their mothers, to have higher PPVT–III and social competence scores and fewer attention and behavior problems at age 3; to have a slightly older mother who was less likely to have had her first child before age 19 and more likely to be highly educated and married to child’s father and to have higher cognitive ability scores and lower harsh parenting scores; and to be less likely to live in poverty. In contrast, after propensity score matching, non–Head Start children looked almost exactly the same as Head Start participants.¹⁴ The *t* statistics did not reveal any significant differences between them, except that Head Start participants tended to have higher scores in PPVT–III, WJ–R, and social competence and lower scores in attention problems at age 5, which were also the raw differences in these outcome measures right before regression adjustment after the matching.

Similarly, Appendix Table C shows the descriptive statistics by reference group on selected child demographics and family background variables before the matching. Overall, children in different reference groups tended to have different characteristics than Head Start participants. For example, mothers of Head Start children also tended to be less educated than those who had children in prekindergarten and other center-based care but were more educated than mothers who used parental care for their children. Compared with Head Start participants, children who had prekindergarten, other center-based, and other nonparental care tended to have higher family income while children who had parental care tended to be more similar to Head Start participants. In contrast, there were no significant differences between children in these reference groups and Head Start participants after the propensity score matching (not in table, available from the authors upon request). This evidence suggested that the propensity score matching approach employed in our study could identify a comparable control group for Head Start participants. Therefore, the analyses within the matched sample might be able to substantially reduce biases associated with those observed covariates in estimations of the effects of Head Start (but would not affect biases associated with any unobserved covariates).

In addition, Figure 1 shows the density of propensity scores predicted in the full sample and subgroups of Head Start and each reference category, respectively. As shown in the first graph, when predicted in the full sample by Head Start participation status (i.e., participating or not), the propensity scores of children in the reference groups tended to be skewed toward zero and quite similar to each other as well as to the overall distribution of non–Head Start children. In contrast, when the propensity scores were predicted within the subgroups of Head Start and each reference category separately, their distributions were very different for both Head Start participants and nonparticipants in each subgroup. Therefore, as discussed earlier, the findings in both Appendix Table C and Figure 1 demonstrated that the evaluation of Head Start effects relative to other specific care arrangements should be conducted separately for the reference groups in order to further account for the issue of selection bias and to provide more sound estimates.

Effects of Head Start Compared With Any Other Care Arrangement

Table 2 shows the results of the models estimating the effects of Head Start compared with any other child care arrangements right before kindergarten. Model 1 was an OLS regression

¹⁴It is beyond the scope of our study to explore why children with the same characteristics of demographics and family background after matching received different child care arrangements. With regard to Head Start, it is well known that its limited budget has been insufficient for serving all eligible children. It is estimated that only between 40% and 66% of eligible children are able to enroll (Garces et al., 2002; Smolensky & Gootman, 2003; Story, Kaphingst, & French, 2006). Who gets into these limited slots may be due in part to luck and in part to location within cities. Some unobserved factors may affect parents’ child care choices as well, such as motivation, as we discuss later.

that only included child demographics. Child pretreatment scores at age 3 were added in Model 2. Model 3 further included mother and family background covariates. In Model 4, city-fixed effects were added to control for the contextual heterogeneity across cities. Model 5 employed a propensity score matching method, as detailed earlier, to further address the issue of selection bias. In all analyses, the outcome variables were standardized with a mean of 0 and a standard deviation of 1. Therefore, the coefficients reported here may be interpreted as effect sizes in terms of changes corresponding to the proportion of a standard deviation (*SD*). Standard errors of the estimates are also shown in the table, and we will discuss the findings that were statistically significant (at $p < .05$ level).

As presented in Table 2, overall, we found significant and consistent effects of Head Start for improvements in children's cognitive development (measured by PPVT-III and WJ-R Letter-Word Identification) and social competence (measured by ASBI) and for reduction of their attention problems at age 5, and no statistically significant effects on their internalizing or externalizing behavior problems. The results from Model 1 when only child demographics were included showed that compared with nonparticipants, Head Start participants tended to have more externalizing behavior problems—0.09 *SDs*, 95% confidence intervals (CIs) [0.02 to 0.17 *SDs*]—but had no significant differences on other measures at age 5. In contrast, when child pretreatment scores at age 3 were included in Model 2, the difference in behavior problems at age 5 between Head Start participants and nonparticipants was no longer statistically significant. The evidence suggested that to reduce the selection bias in estimating the effects of Head Start, it would be important to control for children's pretreatment outcomes in the models. The absolute values of coefficients from Model 2 were substantially smaller than those in Model 3, which included additional mother and family covariates but were in the same positive or negative directions (except for the coefficients of WJ-R). The fact that the effects of Head Start became larger when additional covariates were added in Model 3 suggested that Model 2 estimates were affected by selection bias associated with these additional observed covariates (but of course, some bias might remain). Results from Models 3 and 4 were strikingly similar, suggesting that the addition of city-fixed effects did not have a major effect on the estimates (although the coefficient of WJ-R became statistically significant). We focus on the findings from Model 4 for discussion.

The results from Model 4 in Table 2 show that children who attended Head Start programs right before kindergarten achieved higher cognitive scores measured by PPVT-III (0.08 *SDs*, 95% CIs [0.01, 0.15 *SDs*]) and WJ-R Letter-Word Identification (0.11 *SDs*, 95% CIs [0.03, 0.18 *SDs*]) compared with their peers who did not attend Head Start programs. Head Start participants also received lower scores in attention problems (−0.11 *SDs*, 95% CIs [−0.19, −0.03 *SDs*]). There were no statistically significant differences in internalizing or externalizing behavior problems between Head Start participants and nonparticipants.

The effects of Head Start from the propensity score matching models, as presented in Model 5 of Table 2, were substantially larger than those from the conventional OLS regressions in Models 1–3 or OLS regressions with city-fixed effects in Model 4. In particular, compared with their matched peers who did not attend Head Start, Head Start participants scored higher on PPVT-III (0.19 *SDs*, 95% CIs [0.07, 0.31 *SDs*]) and WJ-R Letter-Word Identification (0.16 *SDs*, 95% CIs [0.04, 0.28 *SDs*]) scales. Head Start participants also had higher scores in social competence (0.14 *SDs*, 95% CIs [0.01, 0.27 *SDs*]) and lower scores in attention problems (−0.16 *SDs*, 95% CIs [−0.29, −0.03 *SDs*]). These findings confirmed that the propensity score matching approach helped identify children who did not attend Head Start but were comparable to Head Start participants, which considerably reduced selection bias on observed covariates and, as a result, increased the estimated magnitude of Head Start effects.

Effects of Head Start Compared With Specific Types of Care Arrangements

To further examine the effects of Head Start compared with specific types of care arrangements, we conducted analyses within subsamples consisting of children who attended Head Start and children who attended one of the other specific care arrangements—parental care, prekindergarten, other center-based care, or other nonparental care. Table 3 shows the results both from unmatched samples based on the OLS regressions with city-fixed effects (Model 4) and from matched samples based on the propensity score matching models (Model 5) in Table 2. Similar to the findings discussed earlier, the results from unmatched and matched samples were consistent in terms of the coefficients' direction and statistical significance on most outcome measures, but the absolute values of Head Start effects from the matched samples tended to be larger than those from unmatched samples. For the discussion that follows, we focus on the findings from the matched samples that were statistically significant (at $p < .05$ level).

The results shown in Table 3 indicate that the effects of Head Start do differ depending on what the reference group is and in ways that are consistent with prior research on the effects of other types of care. Looking first at the comparison with children who had parental care right before kindergarten, Head Start participants scored considerably better in cognitive development and social competence. Specifically, compared with children in parental care, Head Start participants had higher scores on PPVT-III (0.33 *SDs* from matched sample, 95% CIs [0.18, 0.48 *SDs*]), WJ-R Letter-Word Identification (0.46 *SDs*, 95% CIs [0.33 to 0.60 *SDs*]), and social competence (0.24 *SDs*, 95% CIs [0.09, 0.40 *SDs*]). In contrast, when compared with children who attended prekindergarten programs that tend to be associated with cognitive gains, Head Start children scored higher only in social competence (0.15 *SDs*, 95% CIs [0.01, 0.30 *SDs*]). Compared with children who attended other center-based programs, which have been found to be associated with more behavior problems for some children, Head Start participants were more socially competent (0.17 *SDs*, 95% CIs [0.04, 0.29 *SDs*]) and had fewer attention (−0.18 *SDs*, 95% CIs [−0.32, −0.04 *SDs*]) and externalizing (−0.14 *SDs*, 95% CIs [−0.27, −0.01 *SDs*]) behavior problems. Head Start participants also had higher cognitive scores (0.32 *SDs*, 95% CIs [0.17, 0.46 *SDs*]), on PPVT-III, and 0.41 *SDs*, 95% CIs [0.28 to 0.53 *SDs*] on WJ-R) and fewer attention problems (−0.19 *SDs*, 95% CIs [−0.33, −0.06 *SDs*]) than children who had other nonparental care (a mixed category that generally has not been found to be associated with improvements in learning or behavior).

We also conducted *F* tests for the differences between the groups of child care arrangements. In doing so, we used OLS regressions that included dummy variables for all child care arrangements in the full sample ($n = 2,803$), with the same covariates and city-fixed effects as in Model 4 of Table 2. We first conducted joint tests for significant differences between Head Start and all other care arrangements and then separate tests to compare Head Start with each of the other types of child care.¹⁵ The magnitudes of the differences between Head Start and other specific types of child care, as shown by the regression coefficients of their respective dummy variables, were almost identical to those from the results for the “unmatched” samples of subgroups presented in Table 3. The results of the *F* statistics and significance levels are presented in Appendix Table D. The joint tests showed significant differences between Head Start and other care arrangements in the

¹⁵We also conducted *F* tests for significant differences among the non-Head Start care arrangements. Compared with prekindergarten and other center-based care, parental care showed significant differences in the PPVT-III and WJ-R scales models. Significant differences were also found between prekindergarten and other center-based care in externalizing behavior problems and between prekindergarten and other nonparental care in PPVT-III and WJ-R scales models. There were also significant differences between other center-based care and other nonparental care in their effects on PPVT-III and WJ-R scales. There were no significant differences between parental care and other nonparental care.

models for PPVT–III, WJ–R Letter–Word Identification, and attention problems (at $p < .05$). The results from the separate tests showed that Head Start and parental care significantly differed in their effects on PPVT–III, WJ–R Letter–Word Identification, and social competence. There were also significant differences between Head Start and other center-based care in their effects on attention problems and externalizing behavior problems. Significant differences also existed between Head Start and other nonparental care in the models for PPVT–III, WJ–R Letter–Word Identification, and attention problems.

Moderating Roles of Child Gender and Race/Ethnicity

To investigate whether the effects of Head Start were moderated by child gender or race/ethnicity, we estimated two further groups of models, adding interaction terms between Head Start and these two potential sets of moderators (separately). Appendix Table E presents the results from the matched samples based on the propensity score matching models (Model 5 in Table 2). Analyses using OLS models with city-fixed effects produced similar results.¹⁶

As shown in Appendix Table E, overall there were no statistically significant findings on the interaction terms of Head Start with child gender or race/ethnicity. Therefore, the results overall did not provide evidence that the effects of Head Start on children’s school readiness were moderated by child gender or race/ethnicity. However, our ability to detect significant interactions might have been limited by the small sample size.

Discussion

Using data from the FFCWS, a national longitudinal birth cohort study of a large and diverse sample of predominantly low-income children from medium and large U.S. cities, we examined the links between Head Start attendance and urban children’s school readiness. Overall we found significant effects of Head Start on improvements in cognitive development and social competence and reductions in attention problems at age 5, but no statistically significant effects on children’s internalizing or externalizing behavior problems. These findings were robust to model specifications including controls for an extensive set of child and family characteristics, child ability and problems at age 3, city-fixed effects, and propensity score matching models. Our results also showed that the effects of Head Start depended on what the reference group was (i.e., whether it was parental care, prekindergarten, other center-based care, or other nonparental care). Finally, consistent with the findings in the Head Start Impact Study (ACF, 2005), we did not find evidence that the effects of Head Start on school readiness were moderated by child gender or race/ethnicity.

Our estimates of Head Start effects were roughly comparable to those reported in the Head Start Impact Study. For example, when comparing children who attended Head Start with all other children, we found that the effect sizes for PPVT–III and WJ–R Letter–Word Identification from the propensity score matching models (which essentially were treatment-on-treated [TOT] analyses) were 0.19 and 0.16, respectively; while the effect sizes for social competence and attention problems were 0.14 and -0.16 , respectively. Similarly, the effect sizes from TOT estimates reported in the Head Start Impact Study (ACF, 2005) and the Ludwig and Phillips analyses (2007, 2008, 2008) were between 0.13 and 0.17 (among 3-

¹⁶We also tried to conduct propensity score matching analyses within gender and race/ethnicity subgroups. For gender, the magnitudes of Head Start effects for boys and girls were quite close to each other as well as to the overall effects (i.e., no significant differences were detected); nevertheless, the standard errors were large due to smaller sample sizes in the subgroup analyses. For race/ethnicity, most of the subgroups were too small for propensity score matching analyses (e.g., among Head Start participants, 232 children were non-Hispanic Black, but only 28 were non-Hispanic White, 80 were Hispanic, and 46 were other/biracial), especially given that we were matching children within the same cities. Thus, we only presented the results from the analyses of interactions.

year-old children; not significant in 4-year-old group) on PPVT scale and were 0.34 in 3-year-old group and between 0.26 and 0.32 among 4-year-old children on WJ scales. The effect size on the scale of total behavior problems was -0.16 (among 3-year-old children; not significant in 4-year-old group, although the effects on children's social competencies after 1 year of Head Start participation were not significant in either the 3- or 4-year-old group). It should be noted that the FFCWS and the Head Start Impact Study, as well as many other studies, adopted different versions or items of scales to measure children's developmental outcomes, and thus the findings from these studies may not be directly comparable. For example, the Head Start Impact Study used WJ-III Letter-Word Identification, an updated version based on the Cattell-Horn-Carroll theory of cognitive abilities from WJ-R Letter-Word Identification (Schrank, McGrew, & Woodcock, 2001), the latter of which was adopted in the FFCWS. Moreover, social competence was measured with the Social Competencies Checklist in the Head Start Impact Study and with the ASBI Express subscale in this present study. In addition, the scales of behavior problems were measured by the CBCL/1.5-5 in this study while those in the Head Start Impact Study were measured with other items from the 2000 Family and Child Experiences Survey (FACES; ACF, 2006).

One important contribution of our study is that we further investigated the effects of Head Start in comparison with other specific types of child care arrangements. As discussed, in most previous studies, children who attended Head Start have been compared with children who experienced any other type of care settings. Yet, the distribution of care arrangements among the reference group could vary by study. For example, in the Head Start Impact Study, among children in the control group who did not attend Head Start programs, about 48% received parental care, 35% attended other child care centers, and 18% had other nonparental care (recalculated based on ACF, 2005, pp. 3-6). In our study, the proportions differed:¹⁷ 19% of children who did not attend Head Start received parental care, 29% attended prekindergarten, 43% attended other center-based care, and 9% had other nonparental care. The estimates of Head Start effects might be very different across studies depending on the particular care arrangements that children in the control group received. Moreover, estimates in which Head Start children are compared with all other children miss potential variation in the effects of Head Start as compared with other arrangements. For policy makers who face decisions about funding Head Start versus other programs or about investing in program improvements, understanding that variation is important.

When comparing Head Start to other specific types of child care arrangements, we found that Head Start had substantially larger effects on children's cognitive development when it was compared with parental care (i.e., 0.33 on PPVT-III and 0.46 on WJ-R in the matched models) or other nonparental care (i.e., 0.32 on PPVT-III and 0.41 on WJ-R). These findings were consistent with the study by Hill and colleagues (2002), who found substantial and persistent cognitive benefits of high-quality center-based care provided by the Infant Health and Development Program (IHDP) for children who would otherwise have received parental care or homebased nonparental care. In contrast, Head Start had no significant effects on cognitive development when compared with prekindergarten or other center-based care (similar to the Hill et al. analysis of the IHDP data where the comparison was between those who received the treatment but, in the absence of such a program, would have been in some sort of center-based child care; at that time, pre-K programs were just

¹⁷It should be noted that the distributions of child care arrangements in our analysis sample and in the Head Start Impact Study might not be comparable due to different definitions of care arrangements. In the control group of the Head Start Impact Study, children who participated in Head Start at any time during the program year were deemed to be Head Start participants; the terms *other care centers* and *other nonparental care* (i.e., relative and nonrelative care) referred to child care arrangements in which children spent at least 5 hr per week; while *parental care* was defined as the absence of these nonparental care arrangements (ACF, 2005). These definitions were different from ours, as detailed earlier.

beginning, and thus no children attended them). The effects of Head Start on children's social competence and behavior also varied depending on the reference group. Head Start tended to increase children's social competence compared with parental care, prekindergarten, and other centerbased care; to reduce attention problems compared with other center-based care and other nonparental care and to reduce externalizing behavior problems compared with other center-based care.

The recently completed follow-up of children in the Head Start Impact Study reported on average impacts when children were in kindergarten and first grade (ACF, 2010). As discussed earlier, in these analyses, children assigned to Head Start were compared with all others, regardless of the other child care arrangements they attended. In general, sustained impacts were not found. Our findings comparing Head Start participants with children who attended other specific types of arrangements might shed light on why this might have been the case. For example, we found that children who attended Head Start and pre-K programs were quite similar on subsequent cognitive outcomes, whereas children who attended Head Start were performing better than those children in noncenter care (i.e., parental and other nonparental care). Therefore, the comparisons between Head Start and other specific types of care arrangements may provide more information on the variation of Head Start effects, and the presence of some larger effects for specific subgroups, than the overall comparison of Head Start participants versus nonparticipants.

As discussed earlier, selection bias has been a critical issue in observational studies on the effects of Head Start. In our study, because children were sampled from hospitals serving disadvantaged communities in large U.S. cities, both Head Start participants and nonparticipants tended to live in disadvantaged families, and thus the differences between them, although still present, were relatively smaller than those in many other studies. For example, at approximately age 5, the proportion of Head Start participants who had mothers with less than a high school education was 25% in our study and 27% in the Early Childhood Longitudinal Program– Kindergarten (ECLS-K; U.S. Department of Education, National Center for Education Statistics [NCES], 2001), which utilized a nationally representative sample of kindergartners, compared with 25% of those who did not attend Head Start in our study and 12% in the ECLS–K. Similarly, among Head Start children, 59% in our study and 49% in the ECLS–K. lived in poverty at age 5; while among children who did not attend Head Start, 40% in our study and only 15% in the ECLS–K. lived in poverty. Therefore, selection bias may be of less concern in our sample than in many other studies. Indeed, we found significant and positive effects of Head Start in the analyses within the full sample after controlling for child and family covariates. We adopted propensity score matching models to further address the selection bias issue and found substantially larger effects of Head Start.

There were several limitations in our study. First, as shown in the Head Start Impact Study, the variation in Head Start participation (e.g., age at first entering) matters. As noted earlier, we could not track exactly when children entered Head Start programs or how many years of Head Start services they had been receiving from the FFCWS 5-year survey (although we did know that virtually no children were in Head Start at the 3-year follow-up). Thus our findings should be interpreted as the average effects of Head Start participation compared with nonparticipation right before school entry, but we could not further investigate the roles of the child's age at entry or the duration of attendance. Second, as in many other child care studies, parents' reports of Head Start and other child care arrangements may not be valid. Children may not have attended Head Start as their parents reported. For example, it was found that among children in the ECLS–K sample who were identified by school or parent reports as having attended Head Start programs that could be located and that responded to verification surveys, approximately 29% of the children did not actually attend those Head

Start programs (NCES, 2001). Such a high rate of overreporting of Head Start participation could lead to considerably underestimation of Head Start effects (Garces et al., 2002; Ludwig & Phillips, 2007, 2008). In addition, our coding of centerbased care was based on parents' reports on children's attendance on a regular basis, using the most regularly attended program as the child's main center-based care arrangement right before kindergarten. Parents might have different definitions of the term *regular basis*, and thus their reports on center-based care might be inconsistent. These variations in parents' reports of child care arrangements should be kept in mind as well when one interprets the findings in our study. Finally, it should be noted that the propensity score matching method adopted in our study is subject to the assumption of ignorable treatment or selection on observables, which requires that all confounding covariates related to treatment status are observed (Dehejia & Wahba, 1999, 2002; Gibson, 2003; Hill et al., 2003; Hill et al., 2005; Rosenbaum & Rubin, 1983). If any important covariates unrelated to the covariates that were already included in the models were omitted in the predictive models, the estimates of Head Start effects could possibly be biased.¹⁸

Despite these limitations, our findings may provide important implications for policy makers who need to decide whether to allocate scarce public funds to Head Start or to other child care or early education programs. Policy makers also need to decide how to allocate funds within Head Start, given that the budget of Head Start is insufficient to serve all eligible children. Our findings suggest that to the extent that improving cognitive development is a policy goal, it would be important in allocating scarce Head Start funds to try to target children who otherwise would receive only parental care or other nonparental care, since the cognitive benefits to these two groups are the largest. The Economic Stimulus Bill, passed in January of 2009, increases the Head Start budget, making it possible to expand access to underserved groups. Another implication, given the low average level of skills of children attending Head Start programs relative to national norms, could be that these children could benefit from program improvements that increase the capacity of Head Start programs to improve children's cognitive skills. Quality improvement in Head Start has been a focus for some time, and such efforts should continue.

However, our results also make clear that Head Start confers more than cognitive benefits. The importance of behavioral aspects of school readiness is increasingly being recognized (e.g., research and reviews by Duncan et al., 2007; Duncan, Ludwig, & Magnuson, 2007; Raver et al., 2009). Thus, the findings on the beneficial effects of Head Start on social competence and attention problems are relevant. It has been proposed to heighten the emphasis of Head Start on reading and other aspects of academic achievement (e.g., President Bush's proposal in 2003; Ludwig & Phillips, 2007, 2008; Waldfogel, 2006). But in doing so, programs should be adapted in ways that build on their existing strengths rather than shifting focus away from the noncognitive domains. In this study, we found significant effects of Head Start for promoting children's school readiness not only in the improvement of cognitive development but also in the increase of social competence and the reduction of attention problems. It would be short-sighted to change the program in ways that resulted in the loss of those benefits. Rather, quality improvements should focus on keeping the best of Head Start while at the same time strengthening those areas where programs could be improved.

¹⁸For example, parents' motivation has been found to be positively associated with children's development, especially academic achievement (Barnard, 2004; Fan, 2001; Fan & Chen, 2001; Hong & Ho, 2005; Kim, 2002). Similar to many other child care studies, the FFCWS did not collect explicit data on parents' motivation. However, our analyses had included many possible predictors, moderators, or mediators of parents' motivation, such as parents' education, employment, and parenting styles; family structure; household income; home environment; and child gender, age, and race/ethnicity (Davis-Kean, 2005; Fan, 2001; Kohl, Lengua, & McMahon, 2000; Zhan, 2006). Therefore, it was unlikely that parents' motivation would significantly bias our estimates independently from the variables included in the models.

References

- Achenbach, TM.; Rescorla, LA. Manual for the ASEBA Preschool Forms and Profiles. Burlington: University of Vermont, Research Center for Children, Youth, and Families; 2000.
- Aughinbaugh A. Does Head Start yield long-term benefits? *Journal of Human Resources* 2001;36:641–665.
- Barnard WM. Parent involvement in elementary school and educational attainment. *Children and Youth Services Review* 2004;26:39–62.
- Barnett WS, Hustedt JT. Head Start's lasting benefits. *Infants and Young Children* 2005;18:16–24.
- Barnett, WS.; Lamy, C.; Jung, K. The effects of state prekindergarten programs on young children's school readiness in five states. New Brunswick, NJ: Rutgers University, National Institute for Early Education Research; 2005.
- 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.
- Berger L, Brooks-Gunn J, Paxson C, Waldfogel J. First-year maternal employment and child outcomes: Differences across racial and ethnic groups. *Children and Youth Services Review* 2008;30:365–387. [PubMed: 20336171]
- Besharov, DJ.; Call, DM. Head Start falls further behind. *New York Times*; 2009 Feb 8. p. WK12
- Blau, DM. The child care problem: An economic analysis. New York, NY: Russell Sage Foundation; 2001.
- Bradley RH. Children's home environments, health, behavior, and intervention efforts: A review using the HOME inventory as a marker measure. *Genetic, Social and General Psychology Monographs* 1993;119:439–490.
- Bradley RH, Corwyn RF. Externalizing problems in fifth grade: Relations with productive activity, maternal sensitivity, and harsh parenting from infancy through middle childhood. *Developmental Psychology* 2007;43:1390–1401. [PubMed: 18020819]
- Campbell FA, Pungello EP, Miller-Johnson S, Burchinal M, Ramey CT. The development of cognitive and academic abilities: Growth curves from an early childhood educational experiment. *Developmental Psychology* 2001;37:231–242. [PubMed: 11269391]
- Clarke-Stewart, A.; Allhusen, VD. What we know about childcare. Cambridge, MA: Harvard University Press; 2005.
- Côté SM, Borge AI, Geoffroy MC, Rutter M, Tremblay RE. Nonmaternal care in infancy and emotional/behavioral difficulties at 4 years old: Moderation by family risk characteristics. *Developmental Psychology* 2008;44:155–168. [PubMed: 18194014]
- Currie J. Early childhood intervention programs: What do we know? *Journal of Economic Perspectives* 2001;15:213–238.
- 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.
- Davis-Kean P. The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. *Journal of Family Psychology* 2005;19:294–304. [PubMed: 15982107]
- Dehejia RH. Practical propensity score matching: A reply to Smith and Todd. *Journal of Econometrics* 2005;125:355–364.
- Dehejia RH, Wahba S. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association* 1999;94:1053–1062.
- Dehejia RH, Wahba S. Propensity score matching methods for non-experimental 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.
- Duncan GJ, Dowsett CJ, Claessens A, Magnuson K, Huston AC, Klebanov P, Japel C, et al. School readiness and later achievement. *Developmental Psychology* 2007;43:1428–1446. [PubMed: 18020822]
- Duncan GJ, Ludwig J, Magnuson K. Reducing poverty through preschool interventions. *The Future of Children* 2007;17:143–160. [PubMed: 17902264]
- Dunn, LM.; Dunn, LL.; Dunn, DM. Peabody Picture and Vocabulary Test. 3. Circle Pines, MN: American Guidance Service; 1997.
- Fan X. Parental involvement and students' academic achievement: A growth modeling analysis. *Journal of Experimental Education* 2001;70:27–61.
- Fan X, Chen M. Parental involvement and students' academic achievement: A meta-analysis. *Educational Psychology Review* 2001;13:1–22.
- Fragile Families and Child Wellbeing Study. A brief guide to using the mother, father, and couple weights for core telephone surveys Waves 1–4. Princeton, NJ: Princeton University Bendheim-Thoman Center for Research on Child Wellbeing and Columbia University Population Research Center; 2008a.
- Fragile Families and Child Wellbeing Study. Introduction to the Fragile Families public use data: Baseline, one-year, three-year, and five-year core telephone data. Princeton, NJ: Princeton University Bendheim-Thoman Center for Research on Child Wellbeing and Columbia University Population Research Center; 2008b.
- Fuligni AS, Han WJ, Brooks-Gunn J. The infant-toddler HOME in the second and third years of life. *Parenting: Science and Practice* 2004;4:139–159.
- Garces E, Thomas D, Currie J. Long-term effects of Head Start. *American Economic Review* 2002;92:999–1012.
- Gibson CM. Privileging the participants: The importance of sub-group analysis in social welfare evaluations. *American Journal of Evaluation* 2003;24:443–469.
- Gormley W Jr. The effects of Oklahoma's pre-K program on Hispanic children. *Social Science Quarterly* 2008;89:916–936.
- Gormley W Jr, Gayer T, Phillips DA, Dawson B. The effects of universal pre-K on cognitive development. *Developmental Psychology* 2005;41:872–884. [PubMed: 16351334]
- Gormley W Jr, Phillips DA, Gayer T. Preschool programs can boost school readiness. *Science* 2008 June 27;320:1723–1724. [PubMed: 18583597]
- 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 highquality child care. *Journal of Policy Analysis and Management* 2002;21:601–627.
- Hill LJ, Waldfogel J, Brooks-Gunn J, Han W. Maternal employment and child development: A fresh look using newer methods. *Developmental Psychology* 2005;41:833–850. [PubMed: 16351331]
- Hogan AE, Scott KG, Bauer CR. The Adaptive Social Behavior Inventory (ASBI): A new assessment of social competence in high-risk three-year-olds. *Journal of Psychoeducational Assessment* 1992;10:230–239.
- 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.
- Kilburn, MR.; Karoly, LA. The economics of early childhood policy: What the dismal science has to say about investing in children. Santa Monica, CA: RAND Corp; 2008.
- Kim E. The relationship between parental involvement and children's educational achievement in the Korean immigrant family. *Journal of Comparative Family Studies* 2002;33:529–543.

- Kohl GO, Lengua LJ, McMahon RJ. Parental involvement in school: Conceptualizing multiple dimensions and their relations with family and demographic risk factors. *Journal of School Psychology* 2000;38:501–523. [PubMed: 20357900]
- Lanzi RG, Pascoe JM, Keltner B, Ramey SL. Correlates of maternal depressive symptoms in a national Head Start program sample. *Archives of Pediatrics & Adolescent Medicine* 1999;153:801–807. [PubMed: 10437751]
- Lee K. Effects of experimental center-based child care on developmental outcomes of young children living in poverty. *Social Service Review* 2005;79:158–180.
- 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.
- Leuven, E.; Sianesi, B. PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. 2003. Retrieved Feb. 11, 2009, from <http://ideas.repec.org/c/boc/bocode/s432001.html>
- Leventhal T, Martin A, Brooks-Gunn J. The EC-HOME across five national datasets in the third to fifth year of life. *Parenting: Science and Practice* 2004;4:161–188.
- Love JM, Kisker EE, Ross CM, Raikes H, Constantine J, Boller K, Vogel C, et al. The effectiveness of Early Head Start for 3-year-old children and their parents: Lessons for policy and programs. *Developmental Psychology* 2005;41:885–901. [PubMed: 16351335]
- 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.
- Ludwig J, Phillips D. The benefits and costs of Head Start. *Social Policy Report* 2007;21:3–18.
- Ludwig J, Phillips D. Long-term effects of Head Start on low-income children. *Annals of the New York Academy of Sciences* 2008;1136:257–268. [PubMed: 17954676]
- 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]
- Markowitz, J.; Carlson, E.; Frey, W.; Riley, J.; Shimshak, A.; Heinzen, H.; Lee, H., et al. Preschoolers with disabilities, characteristics, services, and results: Wave 1 overview report from the Pre-Elementary Education Longitudinal Study (PEELS). Washington, DC: U.S. Department of Education; 2006.
- Mather, N.; Jaffe, L. Woodcock–Johnson Psycho-Educational Battery- Revised: Reports, recommendations, and strategies. New York, NY: Wiley; 1996.
- Meadows SO, McLanahan SS, Brooks-Gunn J. Parental depression and anxiety and early childhood behavior problems across family types. *Journal of Marriage and Family* 2007;69:1162–1177.
- National Institute of Child Health and Human Development Early Child Care Research Network. Does amount of time spent in child care predict socioemotional adjustment during the transition to kindergarten? *Child Development* 2003;74:976–1005. [PubMed: 12938694]
- National Institute of Child Health and Human Development 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. National Institute of Child Health and Human Development 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
- Raver C, Jones S, Li-Grining C, Zhai F, Metzger M, 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]
- Reichman NE, Teitler JO, Garfinkel I, McLanahan SS. Fragile Families: Sample and design. *Children and Youth Services Review* 2001;23:303–326.
- 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]

- Rigby E, Ryan RM, Brooks-Gunn J. Child care quality in different state policy contexts. *Journal of Policy Analysis and Management* 2007;26:887–907.
- 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.
- Royston P. Multiple imputation of missing values: Update of ICE. *Stata Journal* 2005;5:527–536.
- Rubin, DB. Multiple imputation for non-response in surveys. New York, NY: Wiley; 1987.
- Rutter M. Proceeding from observed correlation to causal inference: The use of natural experiments. *Perspectives on Psychological Science* 2007;2:377–395.
- Schneider, B.; Carnoy, M.; Kilpatrick, J.; Schmidt, WH.; Shavelson, RJ. Estimating causal effects using experimental and observational designs. Washington, DC: American Educational Research Association; 2007.
- Schrank, FA.; McGrew, KS.; Woodcock, RW. Technical abstract. Itasca, IL: Riverside; 2001. Woodcock–Johnson III assessment service bulletin No 2.
- Schweinhart, LJ.; Montie, J.; Xiang, Z.; Barnett, WS.; Belfield, CR.; Nores, M. Lifetime effects: The High/Scope Perry Preschool Study through age 40. Ypsilanti, MI: High/Scope Press; 2005.
- Shonkoff, JP.; Philips, DA., editors. From neurons to neighborhoods: The science of early childhood development. Washington, DC: National Academy Press; 2000.
- Smith JA, Todd PE. Does matching overcome LaLonde’s critique of non-experimental estimates? *Journal of Econometrics* 2005;125:305–353.
- Smolensky, E.; Gootman, JA., editors. Working families and growing kids: Caring for children and adolescents. Washington, DC: National Academies Press; 2003.
- Story M, Kaphingst KM, French S. The role of schools in obesity prevention. *The Future of Children* 2006;16:109–142. [PubMed: 16532661]
- Styfco, S.; Zigler, E. The Head Start debates. Baltimore, MD: Brookes Publishing; 2004.
- U.S. Census Bureau. Statistical abstract of the United States: 2009. 128. Washington, DC: Author; 2008.
- U.S. Department of Education, National Center for Education Statistics.. ECLS–K base year restricted-use Head Start files: CD-ROM and user’s manual (NCES 2001-025). Washington, DC: Author; 2001.
- U.S. Department of Health and Human Services, Administration for Children and Families.. Head Start Impact Study: First-year findings. Washington, DC: Author; 2005.
- U.S. Department of Health and Human Services, Administration for Children and Families.. FACES findings: New research on Head Start outcomes and program quality. Washington, DC: Author; 2006.
- U.S. Department of Health and Human Services, Administration for Children and Families. Program description and program services. 2009. Retrieved on Jan. 20, 2009, from <http://www.acf.hhs.gov/programs/ohs/index.html>
- U.S. Department of Health and Human Services, Administration for Children and Families.. Head Start Impact Study: Final report. Washington, DC: Author; 2010.
- van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine* 1999;18:681–694. [PubMed: 10204197]
- Votruba-Drzal E, Coley RL, Chase-Lansdale PL. Child care and low-income children’s development: Direct and moderated effects. *Child Development* 2004;75:296–312. [PubMed: 15015691]
- Waldfoegel, J. What children need. Cambridge, MA: Harvard University Press; 2006.
- Wechsler, D. Wechsler Adult Intelligence Scale–Revised (WAIS-R manual). San Antonio, TX: Psychological Corp; 1981.
- Yeung WJ, Linver MR, Brooks-Gunn J. How money matters for young children’s development: Parental investment and family processes. *Child Development* 2002;73:1861–1879. [PubMed: 12487499]
- Zhan M. Assets, parental expectations and involvement, and children’s educational performance. *Children and Youth Services Review* 2006;28:961–975.

Appendix

Table A
Selected Descriptive Statistics of Children in and out of the Analysis Sample

Variable	Head Start				Non-Head Start			
	In analysis(<i>n</i> = 386)		Out of analysis(<i>n</i> = 71)		In analysis(<i>n</i> = 2,417)		Out of analysis(<i>n</i> = 1,368)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Care arrangement right before kindergarten								
Head Start	1.00	0.00	1.00	0.00	—	—	—	—
Parental care	—	—	—	—	0.19	0.39	0.16	0.37
Prekindergarten	—	—	—	—	0.29	0.45	0.32	0.47
Other center-based care	—	—	—	—	0.43	0.50	0.43	0.50
Other nonparental care	—	—	—	—	0.09	0.29	0.09	0.28
Child characteristics								
Child male	0.49	0.50	0.63*	0.49	0.53	0.50	0.53	0.50
Age at in-home assessment (months)	64.13	2.52	64.31	0.00	64.17	2.54	64.32*	0.68
Race/ethnicity								
Non-Hispanic White	0.07	0.26	0.07	0.26	0.18	0.39	0.17	0.38
Non-Hispanic Black	0.60	0.49	0.55	0.50	0.47	0.50	0.39**	0.49
Hispanic	0.21	0.41	0.20	0.40	0.20	0.40	0.22	0.41
Biracial/other	0.11	0.32	0.18	0.39	0.15	0.36	0.21**	0.41
Behavioral & cognitive scores at child age 3								
PPVT-III scores	85.23	17.11	83.65	14.16	87.93	17.21	83.45**	15.75
Social competence	15.36	2.40	15.23	2.56	15.62	2.48	15.17**	2.81
Attention problems	2.51	1.51	2.81	1.36	2.30	1.53	2.77**	1.50
Internalizing behavior problems	10.89	5.51	11.18 [†]	5.57	9.60	5.37	9.08**	6.20
Externalizing behavior problems	13.55	7.88	14.01	7.23	12.20	7.42	14.47**	7.79
Mother and household characteristics								
Employment at child's age 3								
Work paid < 35 hrs per week	0.17	0.38	0.15	0.36	0.16	0.37	0.15	0.36
Work paid ≥35 + hrs per week	0.44	0.50	0.44	0.50	0.45	0.50	0.41 [†]	0.49
Looking for job	0.25	0.43	0.28	0.45	0.21	0.40	0.21	0.41
Not working	0.15	0.35	0.13	0.34	0.19	0.39	0.23*	0.42
Education								
Less than high school	0.29	0.45	0.27	0.45	0.27	0.44	0.30*	0.46
High school diploma/GED	0.38	0.49	0.35	0.48	0.28	0.45	0.29	0.45
Some college/tech school	0.29	0.46	0.34	0.48	0.32	0.47	0.27**	0.45
Bachelor's degree/graduate	0.04	0.19	0.04	0.20	0.13	0.34	0.14	0.34
Mother cognitive ability score	6.44	2.29	7.29**	1.69	6.89	2.58	6.82	2.15
Mother depressed in past year	0.24	0.43	0.16	0.37	0.21	0.40	0.21	0.41
Parenting and learning environment								
Harsh parenting score	0.61	1.14	4.59	0.38	0.47	1.03	4.51	0.52
Maternal responsiveness score	5.08	1.09	5.16	0.80	5.14	0.96	5.12	0.63

Variable	Head Start				Non-Head Start			
	In analysis(<i>n</i> = 386)		Out of analysis(<i>n</i> = 71)		In analysis(<i>n</i> = 2,417)		Out of analysis(<i>n</i> = 1,368)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Cognitively stimulating materials	9.44	1.19	9.60	0.64	9.55	1.13	9.45**	0.93
Household income at child's age 3								
Below 50% poverty line	0.31	0.46	0.16*	0.37	0.22	0.41	0.21	0.41
50–100% poverty line	0.27	0.44	0.22	0.42	0.18	0.38	0.18	0.39
100–200% poverty line	0.26	0.44	0.30	0.46	0.25	0.44	0.22*	0.41
200–300% poverty line	0.10	0.30	0.17 [†]	0.38	0.14	0.35	0.15	0.35
300% + poverty line	0.06	0.24	0.14	0.35*	0.21	0.40	0.24*	0.43

Note. Two-tailed *t* statistics testing the mean differences between those Head Start children in the analysis sample and those out of the analysis sample as well as the mean differences between non-Head Start children in the analyses and those out of the analyses. PPVT-III = Peabody Picture Vocabulary Test (3rd ed.; Dunn, Dunn, & Dunn, 1997).

[†] *p* < .10.

* *p* < .05.

** *p* < .01.

Table B

Percentage of Head Start Children in Our Analysis Sample and Head Start Impact Study

Variable	Head Start children in analysis sample(<i>n</i> =386)	Head Start children in Head Start Impact Study(<i>n</i> = 1,530)
Child male	48.7	47.9
Child race/ethnicity		
Non-Hispanic White	7.3	24.3
Non-Hispanic Black	60.4	33.3
Hispanic	20.8	37.0
Other	11.5	5.4
Child's parents were married	19.0	43.7
Biological mother was a teen mom	40.0	36.2
Mother's/primary caregiver's age	27.0	29.5
Mother's education high school or less	66.8	67.1
Biological mother was employed	60.8	51.4
Mother/parents self-reported health as excellent or good	85.1	85.5
Household income < \$1,500/month	55.1	63.1

Note. These selected child demographics and family background variables in our analysis sample (*n* = 386) were based on the Fragile Families and Child Wellbeing Study 3-year survey conducted when children on average were 3 years old, while children in the Head Start Impact Study (*n* = 1,530) were from the 3-year-old treatment group of the study (U.S. Department of Health and Human Services, Administration for Children and Families, 2005).

Table C

Selected Descriptive Statistics by Reference Group

Variable	Head Start (n= 386)		Parental care (n= 451)		Prekindergarten (n= 701)		Other centers (n= 1,041)		Other nonparental (n= 224)	
	M	SD	M	SD	M	SD	M	SD	M	SD
Child characteristics										
Child male	0.49	0.50	0.55 [†]	0.50	0.54	0.50	0.52	0.50	0.54	0.50
Age at in-home assessment (months)	64.13	2.52	64.00	2.56	64.10	2.45	64.23	2.54	64.47	2.75
Race/ethnicity										
Non-Hispanic White	0.07	0.26	0.19 ^{**}	0.39	0.15 ^{**}	0.35	0.21 ^{**}	0.41	0.17 ^{**}	0.37
Non-Hispanic Black	0.60	0.49	0.41 ^{**}	0.49	0.50 ^{**}	0.50	0.49 ^{**}	0.50	0.43 ^{**}	0.50
Hispanic	0.21	0.41	0.25	0.43	0.22	0.42	0.15 ^{**}	0.36	0.23	0.42
Biracial/other	0.11	0.32	0.15	0.36	0.13	0.34	0.15 [†]	0.36	0.18 [*]	0.38
Behavioral & cognitive scores at age 3										
PPVT-III scores	85.23	17.11	84.89	17.44	88.00 ^{**}	16.57	89.14 ^{**}	17.50	88.26 [*]	16.65
Social competence	15.36	2.40	15.17	2.71	15.69 [*]	2.44	15.76 ^{**}	2.35	15.65	2.57
Attention problems	2.51	1.51	2.25 [*]	1.54	2.35 [†]	1.59	2.29 [*]	1.52	2.30 [†]	1.40
Internalizing behavior problems	10.89	5.51	10.06 [*]	5.84	9.24 ^{**}	5.39	9.64 ^{**}	5.34	9.64 ^{**}	5.11
Externalizing behavior problems	13.55	7.88	11.95 ^{**}	7.39	12.20 ^{**}	7.58	12.15 ^{**}	7.33	12.86	7.41
Mother and household characteristics										
Employment at child's age 3										
Work paid < 35 hrs per week	0.17	0.38	0.12 [†]	0.33	0.15	0.36	0.19	0.39	0.13	0.34
Work paid ≥ 35 + hrs per week	0.44	0.50	0.28 ^{**}	0.45	0.52 ^{**}	0.50	0.45	0.50	0.55 ^{**}	0.50
Looking for job	0.25	0.43	0.26	0.44	0.19 [*]	0.39	0.19 [*]	0.39	0.25	0.43
Not working	0.15	0.35	0.34 ^{**}	0.47	0.14	0.35	0.18	0.38	0.07 ^{**}	0.25
Education										
Less than high school	0.29	0.45	0.44 ^{**}	0.50	0.24 [†]	0.42	0.21 ^{**}	0.41	0.30	0.46
High school diploma/GED	0.38	0.49	0.24 ^{**}	0.42	0.28 ^{**}	0.45	0.29 ^{**}	0.46	0.33	0.47
Some college/tech school	0.29	0.46	0.25	0.44	0.35 [*]	0.48	0.33	0.47	0.31	0.46
Bachelor's degree/graduate	0.04	0.19	0.07 [*]	0.26	0.13 ^{**}	0.34	0.17 ^{**}	0.38	0.06	0.24
Mother cognitive ability score	6.44	2.29	6.53	2.67	6.93 ^{**}	2.60	7.01 ^{**}	2.53	6.90 [*]	2.54
Mother depressed in past year	0.24	0.43	0.20	0.40	0.18 [*]	0.39	0.21	0.41	0.27	0.45
Parenting and learning environment										

Variable	Head Start(<i>n</i> = 386)		Parental care(<i>n</i> = 451)		Prekindergarten(<i>n</i> = 701)		Other centers(<i>n</i> = 1,041)		Other nonparental(<i>n</i> = 224)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Harsh parenting score	0.61	1.14	0.45 [†]	1.06	0.49	0.99	0.45 [†]	1.03	0.57	1.14
Maternal responsiveness score	5.08	1.09	5.07	1.00	5.13	0.99	5.18 [†]	0.87	5.09	1.18
Cognitively stimulating materials	9.44	1.19	9.43	1.18	9.54	1.14	9.63 ^{**}	1.09	9.49	1.21
Household income at child's age 3										
Below 50% poverty line	0.31	0.46	0.30	0.46	0.20 ^{**}	0.40	0.19 ^{**}	0.39	0.24 [†]	0.43
50–100% poverty line	0.27	0.44	0.22 [†]	0.41	0.17 ^{**}	0.37	0.16 ^{**}	0.37	0.18 [*]	0.39
100–200% poverty line	0.26	0.44	0.26	0.44	0.27	0.45	0.23	0.42	0.29	0.45
200–300% poverty line	0.10	0.30	0.12	0.33	0.16 ^{**}	0.37	0.14 [*]	0.35	0.15 [†]	0.36
300% + poverty line	0.06	0.24	0.10	0.29 [†]	0.20 ^{**}	0.40	0.27 ^{**}	0.44	0.14 ^{**}	0.34

Note. Two-tailed *t* statistics testing mean differences between Head Start participants and children in each reference group. PPVT–III = Peabody Picture Vocabulary Test (3rd ed.; Dunn, Dunn, & Dunn, 1997).

[†] *p* < .10.

* *p* < .05.

** *p* < .01.

Table D
Significant Tests Between Head Start and Other Specific Care Arrangements (F Statistics With Significance Levels)

Variable	PPVT-III	WJ-R	Social competence	Attention problems	Internalizing problems	Externalizing problems
Joint test: Head Start vs. all other care	17.29**	22.86**	2.38 [†]	2.92*	0.87	2.52 [†]
Separate tests: Head Start vs. other care						
Parental care	14.50**	63.08**	4.63*	2.80	0.26	0.34
Prekindergarten	0.13	0.15	3.58 [†]	2.27	0.35	0.02
Other center-based care	0.45	0.13	3.67 [†]	9.19**	1.41	4.29*
Other nonparental care	21.84**	36.32**	0.34	4.78*	0.19	2.79

Note. *F* tests were conducted after ordinary least squares regressions that included the dummy variables for all child care arrangements in the full sample ($n = 2,803$), in which the same covariates and city-fixed effects as in Model 4 of Table 2 were used. PPVT-III = Peabody Picture Vocabulary Test (3rd ed.; PPVT-III; Dunn, Dunn, & Dunn, 1997); WJ-R = Woodcock-Johnson Psycho-Educational Battery-Revised Letter-Word Identification (Mather & Jaffe, 1996).

[†] $p < .10$.

* $p < .05$.

** $p < .01$.

Table E

Moderating Roles of Child Gender and Race/Ethnicity

Variable	PPVT-III		WJ-R		Social competence		Attention problems		Internalizing problems		Externalizing problems	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Interactions with child gender												
Head Start	0.18 [†]	0.10	0.10	0.11	0.15	0.11	-0.12	0.12	-0.03	0.10	0.07	0.11
Child male	-0.83	0.61	-0.41	0.79	-0.00	0.17	0.05	1.01	-0.35	0.93	-0.67	0.87
Head Start × Child Male	0.02	0.17	0.12	0.19	-0.19	0.20	-0.08	0.20	-0.03	0.16	-0.31	0.19
Interactions with child race/ethnicity												
Head Start	0.09	0.28	0.48	0.30	-0.07	0.21	-0.22	0.30	0.40	0.40	0.33	0.33
Child race/ethnicity ^a												
Non-Hispanic Black	1.90	1.81	0.97	2.20	0.26	0.52	0.18	2.82	0.03	0.57	-0.12	0.28
Hispanic	2.16	1.85	0.65	2.33	0.21	0.36	0.98	2.95	0.51	0.61	0.34	0.34
Biracial/other	0.92	0.97	0.59	1.22	0.40	0.36	0.03	1.49	0.34	0.52	0.33	0.32
Interactions with Head Start												
Head Start × Non-Hispanic Black	0.09	0.30	-0.34	0.32	0.27	0.24	0.11	0.32	-0.47	0.42	-0.24	0.37
Head Start × Hispanic	0.22	0.33	-0.48	0.35	0.09	0.26	-0.24	0.36	-0.55	0.44	-0.48	0.39
Head Start × Biracial/Other	0.08	0.34	-0.13	0.42	-0.34	0.35	0.31	0.42	-0.35	0.51	-0.60	0.43

Note. The analyses were conducted based on the propensity score matching models (i.e., Model 5 in Table 2). The outcome variables were standardized to have a mean of 0 and a standard deviation of 1. The sample size was 750. PPVT-III = Peabody Picture Vocabulary Test (3rd ed.; PPVT-III; Dunn, Dunn, & Dunn, 1997); WJ-R = Woodcock-Johnson Psycho-Educational Battery-Revised Letter-Word Identification (Mather & Jaffe, 1996).

^aNon-Hispanic White omitted.

[†] $p < .10$.

* $p < .05$.

** $p < .01$.

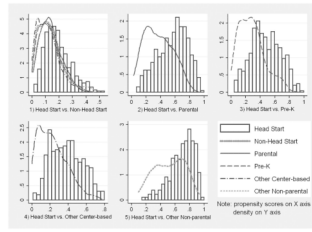


Figure 1. Density of propensity scores predicted in full sample and subgroups of Head Start and each reference category.

Table 1

Descriptive Statistics by Head Start Participation

Variable	Full sample (n= 2,803)		Non-Head Startbefore matching (n= 2,417)		Head Start(n= 386)		Non-Head Start after matching (n= 375)	
	M	SD	M	SD	M	SD	M	SD
Care arrangement right before kindergarten								
Head Start	0.14	0.34	—	—	1.00	0.00	—	—
Parental care	0.16	0.37	0.19	0.39	—	—	0.24	(0.43)
Prekindergarten	0.25	0.43	0.29	0.45			0.27	(0.45)
Other center-based care	0.37	0.48	0.43	0.50			0.38	(0.49)
Other nonparental care	0.08	0.27	0.09	0.29			0.11	(0.35)
Child characteristics								
Child male	0.53	0.50	0.53 [†]	0.50	0.49	0.50	0.52	(0.50)
Age at in-home assessment (months)	64.17	2.54	64.17	2.54	64.13	2.52	63.96	(2.56)
Race/ethnicity								
Non-Hispanic White	0.17	0.37	0.18**	0.39	0.07	0.26	0.07	(0.26)
Non-Hispanic Black	0.49	0.50	0.47**	0.50	0.60	0.49	0.60	(0.49)
Hispanic	0.20	0.40	0.20	0.40	0.21	0.41	0.22	(0.41)
Biracial/other	0.14	0.35	0.15 [†]	0.36	0.11	0.32	0.11	(0.32)
Mother's first child	0.39	0.49	0.40**	0.49	0.32	0.47	0.35	(0.48)
Low birth-weight	0.10	0.30	0.10	0.30	0.11	0.31	0.11	(0.32)
Mother reported fair/poor health	0.03	0.16	0.02	0.16	0.03	0.17	0.03	(0.17)
Behavioral & cognitive scores at age 3								
PPVT-III scores	87.56	17.22	87.93**	17.21	85.23	17.11	85.78	(15.76)
Social competence	15.58	2.47	15.62 [†]	2.48	15.36	2.40	15.34	(2.75)
Attention problems	2.33	1.53	2.30**	1.53	2.51	1.51	2.56	(1.56)
Internalizing behavior problems	9.78	5.41	9.60**	5.37	10.89	5.51	10.50	(6.32)
Externalizing behavior problems	12.40	7.52	12.20**	7.42	13.55	7.88	12.99	(9.40)
Mother & household characteristics								
Age	28.12	5.91	28.30**	5.96	27.02	5.46	27.32	(5.54)

Variable	Full sample (n = 2,803)		Non-Head Start before matching (n = 2,417)		Head Start (n = 386)		Non-Head Start after matching (n = 375)	
	M	SD	M	SD	M	SD	M	SD
First child born when < 19 years old	0.32	0.46	0.30**	0.46	0.40	0.49	0.39	(0.49)
Relationship with father at child's age 3								
Married	0.30	0.46	0.32**	0.47	0.19	0.39	0.19	0.39
Cohabiting	0.20	0.40	0.20	0.40	0.20	0.40	0.20	0.40
Friends/visiting	0.24	0.43	0.23**	0.42	0.30	0.46	0.33	0.47
Other	0.25	0.44	0.24**	0.43	0.31	0.46	0.28	0.45
Employment at child's age 3								
Work paid < 35 hrs per week	0.16	0.37	0.16	0.37	0.17	0.38	0.17	0.38
Work paid ≥ 35 + hrs per week	0.45	0.50	0.45	0.50	0.44	0.50	0.40	0.49
Looking for job	0.21	0.41	0.21†	0.40	0.25	0.43	0.27	0.44
Not working	0.18	0.38	0.19†	0.39	0.15	0.35	0.16	0.37
Education								
Less than high school	0.27	0.44	0.27	0.44	0.29	0.45	0.34	0.47
High school diploma/GED	0.29	0.46	0.28	0.45**	0.38	0.49	0.32	0.47
Some college/tech school	0.32	0.46	0.32	0.47	0.29	0.46	0.29	0.46
Bachelor's degree/graduate	0.12	0.33	0.13**	0.34	0.04	0.19	0.04	0.20
Mother cognitive ability score	6.83	2.55	6.89**	2.58	6.44	2.29	6.59	2.50
Mother depressed in past year	0.21	0.41	0.21	0.40	0.24	0.43	0.23	0.42
Parenting and learning environment								
Harsh parenting score	0.50	1.05	0.47†	1.03	0.61	1.14	0.61	1.13
Maternal responsiveness score	5.13	0.98	5.14	0.96	5.08	1.09	5.15	1.07
Cognitively stimulating materials	9.54	1.14	9.55†	1.13	9.44	1.19	9.45	1.20
Household income at child's birth								
Below 50% poverty line	0.19	0.39	0.17**	0.38	0.27	0.44	0.28	0.45
50–100% poverty line	0.17	0.38	0.16*	0.37	0.21	0.41	0.22	0.42
100–200% poverty line	0.25	0.44	0.25	0.43	0.26	0.44	0.27	0.44
200–300% poverty line	0.15	0.36	0.15	0.36	0.16	0.37	0.15	0.36
300% + poverty line	0.23	0.42	0.26**	0.44	0.10	0.29	0.08	0.27

Variable	Full sample (n= 2,803)		Non-Head Startbefore matching (n= 2,417)		Head Start(n= 386)		Non-Head Start after matching (n= 375)	
	M	SD	M	SD	M	SD	M	SD
Household income at child's age 3								
Below 50% poverty line	0.23	0.42	0.22**	0.41	0.31	0.46	0.32	0.47
50–100% poverty line	0.19	0.39	0.18**	0.38	0.27	0.44	0.25	0.43
100–200% poverty line	0.25	0.44	0.25	0.44	0.26	0.44	0.26	0.44
200–300% poverty line	0.14	0.35	0.14*	0.35	0.10	0.30	0.09	0.29
300% + poverty line	0.19	0.39	0.21**	0.40	0.06	0.24	0.08	0.27

Note. Two-tailed *t* statistics testing mean differences between Head Start participants and nonparticipants before and after propensity score matching. PPVT-III = Peabody Picture Vocabulary Test (3rd ed.; Dunn, Dunn, & Dunn, 1997).

† $p < .10$.

* $p < .05$.

** $p < .01$.

Table 2

Effects of Head Start Compared With Any Other Care Arrangements

Model	PPVT-III		WJ-R		Social competence		Attention problems		Internalizing problems		Externalizing problems	
	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE
Model 1	-0.03	0.04	0.00	0.04	0.06	0.04	-0.02	0.04	0.07 [†]	0.04	0.09*	0.04
Model 2	0.03	0.05	-0.01	0.05	0.09 [†]	0.05	-0.05	0.05	0.00	0.05	-0.01	0.05
Model 3	0.10*	0.05	0.09	0.05	0.11*	0.05	-0.10*	0.05	-0.02	0.05	-0.05	0.05
Model 4	0.08*	0.03	0.11**	0.03	0.10 [†]	0.06	-0.11*	0.04	-0.02	0.03	-0.04	0.03
Model 5	0.19**	0.06	0.16**	0.06	0.14*	0.05	-0.16*	0.07	-0.05	0.06	-0.08	0.06

Note. Model 1 included only child demographics; additional child pretreatment scores at age 3 were added in Model 2; Model 3 contained additional mother and family covariates; Model 4 controlled for city fixed effects; and Model 5 used propensity score matching. The outcome variables were standardized to have a mean of 0 and a standard deviation of 1. The sample sizes were 2,803 for Models 1–4 and 750 for Model 5. PPVT-III = Peabody Picture Vocabulary Test (3rd ed.; Dunn, Dunn, & Dunn, 1997); WJ-R = Woodcock-Johnson Psycho-Educational Battery-Revised Letter-Word Identification (Mather & Jaffe, 1996).

[†] $p < .10$.

* $p < .05$.

** $p < .01$.

Table 3

Effects of Head Start Compared With Other Specific Care Arrangements

Reference care	PPVT-III		WJ-R		Social competence		Attention problems		Internalizing problems		Externalizing problems													
	Unmatched	Matched	Unmatched	Matched	Unmatched	Matched	Unmatched	Matched	Unmatched	Matched	Unmatched	Matched												
	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE												
Parental care	0.21**	0.06	0.33**	0.07	0.39**	0.06	0.46**	0.07	0.17*	0.07	0.24**	0.08	0.08	0.08	-0.02	0.06	-0.07	0.08						
Prekindergarten	0.06	0.06	0.09	0.07	-0.02	0.07	0.01	0.07	0.13*	0.06	0.15*	0.07	-0.06	0.05	-0.05	0.04	0.04	0.03	0.07					
Other center-based	0.04	0.05	0.09	0.07	0.03	0.05	0.05	0.07	0.09 [†]	0.05	0.17*	0.07	-0.15**	0.04	-0.18*	0.07	-0.09*	0.04	-0.14*	0.07				
Other non-parental	0.21**	0.05	0.32**	0.07	0.32**	0.07	0.41**	0.06	0.03	0.08	0.05	0.07	-0.18*	0.08	-0.19**	0.07	-0.01	0.07	-0.18	0.11	-0.13	0.08	-0.10	0.08

Note. The analyses were conducted separately in subsamples consisting of Head Start participants and children who received one of other specific care arrangements. Results from unmatched least squares regressions with city-fixed effects (i.e., Model 4 in Table 2), and results from matched samples were from propensity score matching analyses (i.e., Model 5 in Table 2). The sizes for unmatched and matched samples were 837 and 682 for parental care; 1,087 and 678 for prekindergarten; 1,427 and 732 for other center-based care; and 610 and 596 for other nonparental care, respectively. The outcome variables were standardized to have a mean of 0 and a standard deviation of 1. PPVT-III = Peabody Picture Vocabulary Test (3rd ed.; Dunn, Dunn, & Dunn, 1997); WJ-R = Woodcock-Johnson Psycho-Educational Battery-Revised Letter-Word Identification (Mather & Jaffe, 1996).

[†] $p < .10$.

* $p < .05$.

** $p < .01$.