


Beyond Just the Average Effect: Variation in Head Start Treatment Effects on Parenting Behavior

Christina M. Padilla 
Georgetown University

Parent engagement has been a cornerstone of Head Start since its inception in 1965. Prior studies have found evidence for small to moderate impacts of Head Start on parenting behaviors but have not considered the possibility that individual Head Start programs might vary meaningfully in their effectiveness at improving parenting outcomes. The present study uses the Head Start Impact Study to examine the average effect of random assignment to and participation in Head Start on parenting outcomes as well as variation in that effect across Head Start programs. Findings reveal that Head Start is effective on average at promoting parents' daily reading and overall literacy and math activities with children but that effects vary significantly for parents' literacy and math activities, with some programs much more and some much less effective than their local alternatives. Findings also demonstrate that Head Start has consistent near-zero impacts across centers on parents' disciplinary interactions with children.

Keywords: *early childhood, econometric analysis, educational policy, evaluation, hierarchical linear modeling, multisite studies, parents and families, policy analysis, program evaluation, secondary data analysis*

CHILDREN'S early experiences with their caregivers are one of the most important inputs to their skill development (Crosnoe et al., 2010; Price, 2010; Villena-Roldán & Ríos-Aguilar, 2012; Waldfogel & Washbrook, 2011). Public programs—including home visiting and other parenting interventions—can be effective at changing parents' behaviors (see Ryan & Padilla, 2018, for a review), but they often struggle to recruit and retain low-income parents (Heath et al., 2018), who often have irregular employment and child care schedules (Chaudry et al., 2012; Enchautegui, 2013; Mytton et al., 2014; Prinz & Sanders, 2007). It is thus judicious to consider additional ways to reach low-income families. The present study considers Head Start—a federally funded school readiness program for low-income children and families—as an important albeit often overlooked avenue for promoting positive parenting behaviors.

As one of the largest sources of federal support specifically dedicated to supporting the school readiness of children from low-income families, Head Start is well positioned to promote positive parenting practices in this population. Based on a belief that boosting children's skills requires attention to multiple aspects of children's development, Head Start adheres to a “whole child” model and provides not only preschool education but also medical, dental, mental health, and nutrition services, as well as supports to engage parents in their children's learning and development (Puma et al., 2010). One way Head Start aims to help parents foster their children's development is by enhancing parents' interactions with children (Puma et al., 2010; Zigler &

Valentine, 1979). Indeed, the newly updated Head Start performance standards specifically emphasize efforts to involve parents in their children's early learning through its Parent, Family, and Community Engagement (PFCE) Framework (U.S. Department of Health and Human Services, Administration for Children and Families, 2018a). Given the program's explicit emphasis on involving parents in their children's early learning, Head Start could be an ideal program through which to support parenting.

Prior studies examining whether Head Start is effective at enhancing parents' interactions with their children have found evidence for small to moderate impacts on the frequency with which parents engage in certain reading and math activities with children and their likelihood of spanking children (Ansari et al., 2016; Gelber & Isen, 2013; Puma et al., 2010). Not all Head Start centers are the same, however. Individual Head Start centers vary considerably with regard to the curricula they use, the parenting programming they offer, their classroom quality, and the ways in which they choose to engage families (Aikens et al., 2017; Moiduddin et al., 2012; U.S. Department of Health and Human Services, Administration for Children and Families, 2016, 2018a). For example, all Head Start centers are required to offer two home visits to families per year, but they can provide more if they choose, producing variability in the number of home visits that are offered across centers (U.S. Department of Health and Human Services, Administration for Children and Families, 2016). Moreover, the PFCE framework makes clear that individual grantees



have freedom in deciding what family engagement at individual centers looks like, given family- and community-specific needs (U.S. Department of Health and Human Services, Administration for Children and Families, 2018a). Prior studies of Head Start's impact on parenting have not considered the likely possibility that individual Head Start programs vary in their effectiveness at improving parenting outcomes. Prior research has thus not been able to fully elucidate the conditions under which Head Start is most effective at reaching its goal of engaging parents in their children's development.

The present study takes advantage of the multisite design of the Head Start Impact Study (HSIS), a large randomized controlled trial (RCT) of Head Start centers, to examine the impact of random assignment to and participation in Head Start on parenting outcomes in two domains—cognitive stimulation and physical discipline. First, I examine Head Start's average effects on parents' cognitively stimulating and disciplinary interactions with their children. Though largely a replication of prior work (Gelber & Isen, 2013; Puma et al., 2010), I focus on parenting measures that are both predictive of children's outcomes and theoretically modifiable by Head Start—parents' daily reading, other literacy activities, math activities, and physical discipline—rather than examining every available variable related to parenting in the HSIS, as in prior studies. The key innovation of this study, however, is the examination not only of average effects of Head Start but also the variation in effects on parents across sites. Solely focusing on the average effect of Head Start could mask meaningful variation in effects across programs that could help to elucidate why some programs are more effective than others (Bloom et al., 2017; Bloom & Weiland, 2015; Weiss et al., 2014; Weiss et al., 2017). It is only by understanding *both* the average effects and variation in those effects that we can form a clear picture of the extent to which Head Start is effective at its goal of improving parents' early interactions with their children.

Head Start as an Intervention for Parents

Since its inception in 1965 as part of the War on Poverty, Head Start has served more than 36 million children and families, making it one of the largest sources of federal support for low-income families with young children in the United States (U.S. Department of Health and Human Services, Administration for Children and Families, 2018b). The program is free for all children whose family income is below the poverty line and, when there are slots available, for an additional 35% of children whose families' income is below 130% of the poverty line (Improving Head Start for School Readiness Act of 2007). Ten percent of Head Start slots are also reserved for children with disabilities, regardless of family income (U.S. Department of Health and Human Services, Administration for Children and Families,

2016). The main goal of Head Start is to improve the school readiness of children from low-income families by supporting the development of the whole child (Office of Head Start, Administration for Children and Families, 2019; U.S. Department of Health and Human Services, Administration for Children and Families, n.d.). To achieve this goal, Head Start provides a number of services, including parent and family supports that address families' self-sufficiency goals, promote positive parent-child relationships, and actively encourage parents' involvement in their children's learning (Office of Head Start, Administration for Children and Families, 2019; Puma et al., 2010; Zigler & Styfco, 2004). The majority of attention surrounding Head Start has centered around the extent to which Head Start supports children through early childhood education (ECE) services. Significantly less attention has been paid to the extent to which Head Start supports parents and families as a whole, despite the known important contributions of parents to children's development as well as Head Start's historical commitment to involving parents in their children's learning.

For more information on which parenting behaviors we might expect Head Start to impact and why we might expect them to be affected, as well as a review of prior research examining Head Start's impact on parenting behavior, see online Supplemental Appendix A.

Variation in Head Start Program Effects on Parenting Behaviors

Prior research suggests that Head Start participation can enhance parenting behavior on average, in terms of both cognitive stimulation and physical discipline (online Supplemental Appendix A). However, not all Head Start centers are the same. Under federal regulations, Head Start centers must all meet certain standards and requirements, but individual centers also have a large degree of freedom in their operation. Individual centers can differ in, for example, their program curricula, the parenting programming they offer, their teachers' qualifications, or even the overall climate and attitude toward parents (Aikens et al., 2017; Bloom & Weiland, 2015; U.S. Department of Health and Human Services, Administration for Children and Families, 2016). It is thus quite possible that centers vary meaningfully in their effectiveness. It is important to note, however, that the opposite could also be true—while Head Start allows individual grantees certain choices, it ultimately operates as a single, uniform program with the same overarching goals and design. It is therefore also possible that Head Start centers do not vary meaningfully in their effectiveness and that the way we have historically thought about the program—as a monolith—has been accurate. Indeed, as detailed below, large multisite RCTs do not always find treatment variation across sites (Weiss et al., 2017). Estimating whether Head Start programs do indeed vary significantly (i.e., above and

beyond random estimation error) in their effectiveness is the first step toward understanding what determines program effectiveness—if some centers are shown to be significantly more effective than others, we can begin to identify practices and features of particularly effective centers and, ideally, implement these practices on a broader scale in order to improve services and outcomes for more children and families.

In recognition of the need to study treatment effect variation, new and innovative methodologies have been developed to precisely quantify treatment effect variation using multisite RCTs (Bloom et al., 2017; Raudenbush & Bloom, 2015; Weiss et al., 2017). Recent studies have used these methodologies to quantify variation in Head Start's effects on children's outcomes (Bloom & Weiland, 2015; Morris et al., 2018; Walters, 2015). They have found that the originally reported small impacts on children's academic and behavioral skills from the HSIS were masking large variation in effects across sites and that programs ranged in their effectiveness at promoting children's school readiness compared to their local alternatives (Bloom & Weiland, 2015; Morris et al., 2018; Walters, 2015). While these studies highlight the potential for Head Start centers' effectiveness to vary, they focused solely on children's outcomes. We have a very limited understanding of the conditions under which Head Start centers are best able to promote parenting behaviors that are known to affect children's school readiness outcomes.

Prior studies have demonstrated that *on average*, Head Start has positive but relatively small effects on the frequency with which parents engage in math and literacy activities and on parents' physical discipline (Ansari et al., 2016; Gelber & Isen, 2013; Puma et al., 2010). Significant average effects do not necessarily signal that treatment effects vary across programs, however, just as null average effects do not mean that there is no variation in impacts. Indeed, in a study that estimated both average effects and effect variation using data from 16 large multisite RCTs of various education and workforce development interventions, Weiss et al. (2017) found that treatment impacts from these 16 studies fell into four categories. First, it is possible to observe consistent near-zero impacts across all sites, wherein the average treatment effect is zero or close to zero and treatment effects do not vary, indicating that the program is consistently ineffective. Second, there could be a near-zero average impact with substantial cross-site variation. In this scenario, average null effects might lead to the conclusion that the program is ineffective, yet the wide variation in effects demonstrates that this is not true for all sites. Third, treatment effects could be consistently positive (or negative), meaning that there is an average effect and that there is little variation in that effect. Fourth, there could be a significant positive impact on average with substantial cross-site variation, where most sites produce positive

effects with widely varying magnitudes. Given the prior findings of small positive average effects of Head Start on parenting outcomes, variation in treatment impacts could mean either that Head Start is consistent in making small changes on parenting behaviors (i.e., no variation in impacts) or that some sites make large changes while others make small or no changes (i.e., significant variation in impacts).

Studying Treatment Effect Variation in the HSIS

The design of the HSIS makes it the ideal data set with which to answer questions about variation in Head Start treatment effects. Specifically, randomization of children and parents to either treatment or control occurred at the center level in the HSIS, meaning that parents applied to a specific Head Start center and then, if randomly assigned to the treatment group, were offered services in that center. This means that the HSIS can be considered a composite of over 300 RCTs, each of which was conducted at a single site. This unique design allows for the estimation of both an average effect and a standard deviation of that effect that quantifies how much the individual effects vary across centers. Very few other RCTs—even federally funded ones—have the power to identify potential effect heterogeneity across program sites. By examining and describing the full range of Head Start effects on parenting behaviors, the present study adds to our knowledge of Head Start's ability to improve parenting practices, a question fundamental to the program's "whole child" model and PFCE framework.

Method

Data and Sample

The HSIS is a nationally representative sample of 383 randomly selected Head Start centers and 4,667 newly entering, eligible 3- and 4-year-old children. The study began in 2002 and continued through 2006, when children ended first grade. In order to maximize statistical power and so that results could be compared to those from prior studies examining treatment effect variation in the HSIS (Bloom & Weiland, 2015; McCoy et al., 2016), the current study combined the 3- and 4-year cohorts and examined short-term parenting outcomes after 1 year of Head Start. Children whose parents applied for Head Start services were randomly assigned to either a treatment group that had access to Head Start or a control group that did not have access to the Head Start center where they applied (but could enroll in other programs).

I created separate analytic samples for each parenting outcome. I restricted each sample to those who had valid data for the parenting outcome under investigation. Following others (Bloom et al., 2017; McCoy et al., 2016), I addressed missing predictor variable data (e.g., missing data on covariates and baseline parenting variables) through

single replication of a multiple imputation model. I imputed using the ICE command in Stata 15.0, which is based on a regression switching protocol using chained equations (Royston, 2005). For each analytic sample, I dropped centers if after restricting to those with nonmissing outcome data they had either no treatment group members or no control group members and thus could not provide experimental estimates of Head Start effects (Bloom & Weiland, 2015; McCoy et al., 2016; across samples, n of children ranged from 3,497 to 3,602 within 316 to 318 centers). I also dropped centers if they had zero compliance with random assignment—meaning that the proportion of their control group members who enrolled in Head Start was equal to the proportion of their treatment group members who enrolled—because these centers similarly could not provide information about Head Start effects. This process resulted in sample sizes ranging from 3,416 to 3,525 children within 296 to 297 Head Start centers.

Measures

Parents answered questions about their interactions with children at the beginning (Fall 2002; Time 1) and end (Spring 2003; Time 2) of the Head Start year in two domains: cognitive stimulation and physical discipline. I examined three cognitive stimulation outcomes: whether parents read to children daily as well as the frequency with which they engaged in multiple literacy- and math-related activities with children, and two physical discipline outcomes: whether parents spanked their children in the past week and the number of times they did so. For more detailed information about each of these measures, as well as the covariates included in all analyses, see online Supplemental Appendix B.

Analytic Strategy

To quantify variation in Head Start's effects on parenting outcomes, I employ recently developed approaches designed for use with multisite RCTs (Bloom et al., 2017; Bloom & Weiland, 2015; Raudenbush & Bloom, 2015; Raudenbush et al., 2012). I separately estimate the causal effect of random assignment to Head Start and the causal effect of enrollment in Head Start.

In all models, parenting behaviors at the end of the Head Start year (Time 2) are predicted by Head Start treatment status and all covariates. Each parenting measure from the beginning of the Head Start year (Time 1) is also included as a covariate, which controls for any potential unmeasured, time-invariant differences in treatment and control group parents present at the beginning of the study (Cain, 1975; Chase-Lansdale et al., 2003; National Institute of Child Health and Human Development Early Child Care Research Network & Duncan, 2003; Votruba-Drzal & Chase-Lansdale, 2004). Regression coefficients are thus interpreted as the effect of

random assignment or enrollment on *changes* in the parenting measure being predicted (Kessler & Greenberg, 1981).

Due to significant noncompliance to random assignment in the HSIS (14%–20% depending on the age cohort and random assignment condition), it was necessary to separately estimate the effect of Head Start *random assignment*—which compares treatment and control group members regardless of their actual participation in the program—and the effect of Head Start *participation*—which adjusts for this noncompliance. Both of these estimates are useful. The effect of random assignment is arguably more policy relevant in that it reflects real-world scenarios in which parents are offered a slot in Head Start but do not always take it, while the effect of program participation may be more relevant to program implementation efforts in that it characterizes the effect for families who actually participated in the program (Puma et al., 2010).

Following others (e.g., Bloom & Weiland, 2015; Morris et al., 2018; Puma et al., 2010), I consider p values less than .10 as statistically significant because this matches the threshold used in the original HSIS final impact report and allows for broader discussion of treatment effect heterogeneity. Results that are significant at the $p < .10$ should be treated with some caution, however.

Effect of Random Assignment. To estimate the effect of random assignment to Head Start—the “intent to treat” (ITT) effect—as well as the cross-site variance in the effect of random assignment, I estimate a two-level random coefficients model that nests parents (Level 1) in the Head Start centers at which they initially sought care (Level 2). First, this two-level random coefficients model specifies a fixed intercept and a random coefficient for each site (called an FIRC approach; Sabol et al., 2020). Specifying a fixed intercept eliminates biases resulting from systematic and nonsystematic differences across Head Start centers (Bloom et al., 2017; Bloom & Weiland, 2015; McCoy et al., 2016), and specifying a random coefficient allows treatment effects to vary across Head Start sites. The model also specifies separate individual-level residual outcome variances for the treatment group and the control group, which accounts for “T/C heteroskedasticity”—the possibility that the individual-level outcome variance for treatment members differs from that of the control group—which could bias the estimate of cross-site variance (Bloom et al., 2017; Bloom & Weiland, 2015). For more details on this approach, see online Supplemental Appendix C.

Effect of Enrollment. In the case of the HSIS, where there are both “no shows” and “crossovers,” the effect of Head Start *enrollment* represents a “local average treatment effect” (LATE), meaning that the average treatment effect is specific to compliers (Angrist et al., 1996; Raudenbush et al., 2012). To estimate the LATE, I use an extension of a typical

instrumental variables approach (Angrist et al., 1996; Heckman & Robb, 1985; Raudenbush et al., 2012) proposed by Raudenbush et al. (2012; “Option C”) and then expanded upon by Bloom and Weiland (2015). This approach allows for estimation of cross-site variation in enrollment effects, accounts for the nested structure of the data, and accounts for random assignment noncompliance. For more details on this approach, see online Supplemental Appendix C.

Results

Descriptive Results

At baseline, treatment and control group members were balanced on almost all covariates (Table 1). The two exceptions were that parents in the treatment group were slightly less likely to have less than a high school education and that the time between baseline and spring parent interviews was slightly longer for control group compared to treatment group parents.

Treatment group parents exhibited slightly more positive parenting behaviors at baseline—they were more cognitively stimulating, read to children more often, read for longer, were less likely to spank, and spanked their children less frequently. These differences were small but nonetheless signal a need to compare changes in parenting across the two groups as a result of Head Start and to control for demographic characteristics of families that might help account for preexisting differences between the groups.

Treatment Impact Average Effects and Variation in Effects

I standardized effect sizes for literacy activities and math activities by dividing the estimated Head Start effect on each outcome in its original units by the control group standard deviation for that outcome (Bloom & Weiland, 2015; Weiss et al., 2017). This is a useful approach in interpreting these outcomes since these scale scores do not have a natural meaning and are thus difficult to interpret (Weiss et al., 2017). For example, if the average literacy activities score for the control group at the end of the Head Start year was 3.5, and the treatment effect on the original 1- to 6-point scale was 0.5, this would mean that on average, control group parents read to their children two or three times per month (scale score of 3) and one or two times per week (scale score of 4), while treatment group parents on average read to children one or two times per week (scale score of 4)—a difference of about two to five times per month. This interpretation gets complicated and tedious, particularly when effects are smaller than 0.5. Putting effect sizes for literacy and math activities into standard deviation units also allows effects to be more easily compared to previous studies. An average effect size for literacy outcomes of 0.2 in standard deviation units would signify that on average, Head

Start parents increased their literacy outcomes more than control group parents by a magnitude that equals 0.2 standard deviations of the literacy activities measure in its original 1- to 6-point scale units.

For consistency, coefficients for spanking frequency are also reported in standard deviation units in tables, but the Results section also notes the interpretation in terms of the item’s natural units (spanking frequency) where significant. Coefficients for binary outcome variables (child is read to daily, child spanked in past week) represent the difference between the treatment and control group in how much parents’ probability of doing that behavior changed as a result of Head Start.

Results from all models examining the average effect of Head Start on parenting outcomes were very similar to results from prior studies examining this effect in the HSIS using similar parenting variables (Gelber & Isen, 2013; Puma et al., 2010). Online Supplemental Appendix D shows the results from the present study for average effects compared to those from prior studies.

ITT Results

Cognitive stimulation. Results for the average effect of random assignment to Head Start on each parenting outcome, as well as the standard deviation for those treatment effects, are displayed in the top panel of Table 2. Random assignment to Head Start was effective on average at promoting parents’ cognitively stimulating interactions with their children. Specifically, parents randomly assigned to Head Start increased their probability of reading to children daily more (by 4.4 percentage points) on average than did parents in the control group. Head Start parents also increased their literacy activities by 0.157 standard deviation units (equivalent to 0.18 points on the original 1- to 6-point scale) and their math activities by 0.179 standard deviation units (equivalent to 0.21 points on the original 1- to 6-point scale) with children more on average than did control group parents.

For two of these outcomes, parents’ literacy and math activities, effects also varied significantly across centers, indicating that some programs were significantly less and some were significantly more effective than their local alternatives. For literacy activities, the significant standard deviation of 0.221 around the average effect of 0.157 indicates that if effect sizes were normally distributed across centers, then 90% of centers would have an effect size for literacy outcomes between -0.21 and 0.52 . This is a wide range and demonstrates the inadequacy of solely considering the average effect. Likewise, for math activities the significant standard deviation of 0.201 around the average effect of 0.179 indicates that if effect sizes were normally distributed across centers, then 90% of centers would have an effect size for math activities between -0.15 and 0.51 —again, a wide range of effects. Effects for the probability of reading daily did not vary significantly across centers.

TABLE 1
Descriptive Statistics and Baseline Balance of the Analytic Sample

Variables	Combined <i>M</i> /% (<i>N</i> = 3,527)	<i>SD</i>	Minimum	Maximum	Control group <i>M</i> /% (<i>N</i> = 1,357)	Treatment group <i>M</i> /% (<i>N</i> = 2,170)	Difference	<i>p</i> of difference
Parenting behaviors at baseline (Fall 2002)								
Cognitive stimulation	1.162	0.374	0	2	1.129	1.183	0.054	.000***
Child is read to daily (%)	0.345	0.475	0	1	0.331	0.354	0.023	.160
Child spanked in previous week (%)	0.455	0.498	0	1	0.478	0.441	-0.037	.032*
Spanking frequency	0.892	1.442	0	15	0.976	0.840	-0.136	.009***
Child characteristics								
Child was in 3-year-old cohort (%)	0.553	0.497	0	1	0.552	0.554	0.002	.909
Female (%)	0.500	0.500	0	1	0.494	0.504	0.009	.595
Black (%)	0.302	0.459	0	1	0.295	0.306	0.011	.479
Hispanic (%)	0.371	0.483	0	1	0.371	0.371	0.000	.999
White (%)	0.327	0.469	0	1	0.334	0.323	-0.011	.490
Home language is Spanish (%)	0.295	0.456	0	1	0.293	0.296	0.003	.835
Family characteristics								
Maternal age (years)	29.170	7.448	16	78	28.977	29.291	0.314	.218
Maternal education (%)								
<High school	0.380	0.485	0	1	0.398	0.368	-0.030	.078 [†]
High school	0.332	0.471	0	1	0.326	0.336	0.009	.560
>High school	0.288	0.453	0	1	0.276	0.296	0.020	.194
Maternal marital status (%)								
Single	0.395	0.489	0	1	0.393	0.396	0.004	.835
Married	0.448	0.497	0	1	0.455	0.444	-0.011	.527
Previously married	0.157	0.364	0	1	0.153	0.160	0.007	.557
Mother is a teenager (%)	0.164	0.370	0	1	0.173	0.158	-0.015	.242
Mother is a recent immigrant (%)	0.186	0.389	0	1	0.191	0.183	-0.008	.558
Child lives with both biological parents (%)	0.498	0.500	0	1	0.502	0.495	-0.006	.709
Assessment characteristics								
Child age at spring testing (years)	4.528	0.567	3.250	6.288	4.535	4.524	-0.011	.571
Child was tested in Spanish (%)	0.249	0.433	0	1	0.243	0.253	0.010	.511
Time between baseline and spring parent interview (weeks)	33.442	3.309	26	48	33.808	33.213	-0.594	.000***

Note. Sample includes children in complete randomized blocks with nonzero compliance. This sample is from a single replicate of a multiple imputation model that imputed missing baseline parenting variables and covariates. The *N* for the descriptive statistics reported here reflects the largest possible of these samples, from the cognitive stimulation variable.
****p* < .001. ***p* < .01. **p* < .05. [†]*p* < .10.

TABLE 2
Average Effects of Head Start Random Assignment and Enrollment on Parenting Outcomes and Standard Deviations of Effects

Statistic	Child is read to daily (y/n)	<i>p</i>	Literacy activities	<i>p</i>	Math activities	<i>p</i>	Child spanked in past week (y/n)	<i>p</i>	Spanking frequency	<i>p</i>
Average effect	0.044	0.005**	0.157	0.000***	0.179	0.000***	-0.021	0.195	-0.016	.644
SD of effect	0.059	0.152	0.221	0.000***	0.201	0.007**	0.034	0.232	0.209	.083 ⁺
Range of effects ^a	(-0.05 to 0.16)		(-0.21 to 0.52)		(-0.15 to 0.51)		(-0.08 to 0.04)		(-0.03 to 0.03)	
			Effects of assignment to Head Start (intent to treat)							
Average effect	0.052	0.008**	0.199	0.000***	0.236	0.000***	-0.037	0.053 [†]	-0.029	.503
SD of effect	0.077	0.195	0.174	0.003**	0.176	0.001**	0.000	0.343	0.264	.110
Range of effects ^a	(-0.07 to 0.18)		(-0.08 to 0.48)		(-0.05 to 0.53)		(-0.04 to -0.04)		(-0.04 to 0.04)	
<i>N</i> children	3,516		3,416		3,462		3,525		3,512	
<i>N</i> centers	297		296		297		297		296	

Note. Models were fit using children with available outcome data in nonzero compliance and complete randomized blocks and using fixed intercepts for centers with a random treatment coefficient. Covariates include those used in the original Head Start Impact Study final report, including child gender, child race/ethnicity, home language, mother's age, mother's education, mother's marital status, whether mother was a teen mother at the time of child's birth, mother immigrant status, whether child lives with both biological parents, and number of weeks between fall and spring parent interviews. Models also include a binary indicator of the child's age cohort. The appropriate parenting measure from baseline is also included as a covariate for each outcome. Effect sizes for all continuous outcomes (literacy activities, math activities, spanking frequency) were calculated by dividing the estimated Head Start effect on each outcome in its original units by the control group standard deviation for that outcome. Coefficients for binary outcome variables (child is read to daily, child spanked in past week) represent the difference between the treatment and control group in how much parents' probability of doing that behavior changed as a result of Head Start. Standard errors are in parentheses.

^aAssuming normality, 90% of cases would fall in this range.

****p* < .001. ***p* < .01. **p* < .05. [†]*p* < .10.

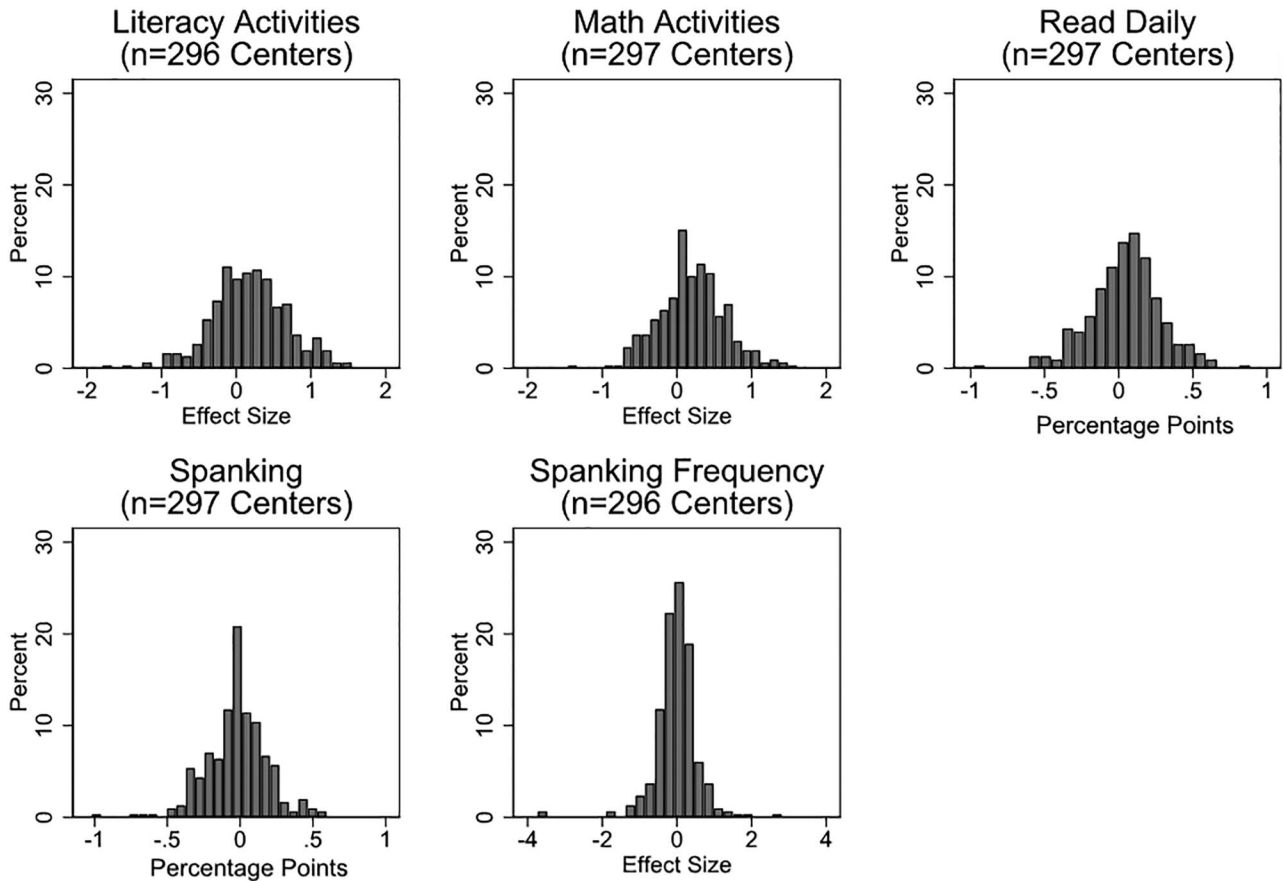


FIGURE 1. Histograms of adjusted empirical Bayes estimates of random treatment effects from FIRC models for each outcome. Effect sizes for continuous outcomes (literacy activities [$SD = 1.16$], math activities [$SD = 1.45$], and spanking frequency [$SD = 1.46$]) are in standard deviation units. For example, an average effect size for literacy (or math) outcomes of 0.2 in standard deviation units would signify that on average, Head Start parents increased their literacy (or math) outcomes more than control group parents by a magnitude that equals 0.2 standard deviations of the literacy (or math) activities measure in its original 1- to 6-point scale units. The original units for spanking frequency was the number of times the parent spanked the child in the past week. Effect sizes for dichotomous outcomes (read daily and spanking) are in percentage point units and represent the difference in treatment and control group parents' likelihood of reading daily and spanking children after 1 year of Head Start. FIRC = fixed intercept and random coefficient.

Figure 1 displays histograms of adjusted empirical Bayes estimates for all examined outcomes from the FIRC models in order to visualize the variation in effects (Bloom & Weiland, 2015). For literacy activities, for example, the figure shows that many programs had an effect size around 0.16 (the average) but that many programs' effect estimates also varied from the average, from negative to positive. The distribution for this effect is wide and relatively low. By contrast, a higher percentage of programs' effects on daily reading, which did not vary significantly, were around the average (0.04, or 4 percentage points). The distribution for this effect is more tightly concentrated around the average.

Figures displaying treatment impact estimates and their standard errors for each outcome for a random sample of centers are included in online Supplemental Appendix E.

Physical discipline. Random assignment to Head Start was not effective on average at lowering the likelihood of parents spanking their children or the frequency with which parents reported spanking their children (Table 2). Although neither of these average effects were significant, this effect did vary—though modestly—for the frequency with which parents spanked children (Figure 1). Specifically, the significant standard deviation of 0.209 around the average effect of -0.016 (equivalent to an average reduced spanking frequency for Head Start parents by 0.02 spans per week) indicates that if effect sizes were normally distributed across centers, then 90% of centers would have effect sizes between -0.03 and 0.03 (-0.5 and 0.5 spans per week) for spanking frequency. This is not a very wide range and all effect sizes are very close to zero. Together, results suggest that Head Start random assignment effects for physical discipline are small and vary minimally.

LATE Results

Cognitive stimulation. Results for the effect of Head Start enrollment were similar to those for the effect of random assignment with regard to the pattern of significant findings, with significant average effects for reading daily, literacy, and math activities, and significant standard deviations around those average effects for literacy and math activities. The magnitude of these effects differed somewhat, however, with larger average effects across all outcomes in the LATE models compared to ITT models (as is common when comparing effects of random assignment versus participation because of “dilution” in ITT effects as a result of noncompliance with random assignment) but smaller standard deviations around those effects in the LATE models compared to ITT models (likely due to increased precision as a result of eliminating uncertainty introduced by noncompliers in LATE estimates). Specifically, the average effect of enrollment for literacy activities was 0.199 (equivalent to 0.23 points on the original 1- to 6-point scale) with a significant standard deviation of 0.174, indicating that if normally distributed, 90% of centers’ effect sizes would be between -0.08 and 0.48 . The average effect for math activities was 0.236 (equivalent to 0.27 points on the original 1- to 6-point scale) with a significant standard deviation of 0.176, indicating that if normally distributed, 90% of centers’ effect sizes would be between -0.05 and 0.53 .

Physical discipline. Whereas random assignment to Head Start did not significantly affect parents’ likelihood of spanking on average, the magnitude of the analogous coefficient on enrollment in Head Start in the LATE model was larger and was statistically significant ($p = .053$), showing that enrollment in head Start reduced the likelihood of parents spanking their children somewhat, by 3.7 percentage points. Enrollment did not affect parents’ average frequency of spanking, however, and neither of these effects varied significantly in LATE models. Results suggest that Head Start enrollment was minorly effective at reducing spanking likelihood on average across programs.

Discussion

One of the hallmarks of Head Start is its focus on empowering parents to be active participants in their children’s learning and development. This study assessed whether Head Start programs are effective at achieving that goal by examining the average effect of Head Start on parents’ cognitively stimulating and disciplinary interactions with their children, as well as the extent to which Head Start programs vary in their effectiveness at improving these interactions. I found that both random assignment to and participation in Head Start were effective on average at improving parents’

cognitively stimulating interactions with children—including their likelihood of reading to children daily and their overall literacy and math activities. For both random assignment and participation, Head Start effects on parents’ literacy and math activities also varied significantly. Effects were much weaker for Head Start’s effects on parents’ physical discipline behavior. Only enrollment—but not random assignment—had a significant effect on decreasing parents’ likelihood of spanking, and this effect was very small in magnitude. Neither random assignment nor enrollment affected parents’ average frequency of spanking. Effects on parents’ likelihood of spanking did not vary significantly across Head Start programs, though the effect of random assignment on spanking frequency varied slightly across programs.

Results from the present study indicating that Head Start is moderately effective at promoting parents’ cognitively stimulating interactions with children are consistent with prior reports (Gelber & Isen, 2013; Puma et al., 2010). This study extends this past work by demonstrating that two of the three examined effects varied significantly across Head Start sites, meaning that some programs were significantly more effective than others at improving parents’ literacy and math activities with children. Moreover, the wide ranges of program effects for both of these outcomes—from negative to positive—demonstrate that while most Head Start programs are more effective than their local alternatives (i.e., the control condition), some programs are *less* effective than their local alternatives at improving the frequency with which parents engage in literacy and math activities with their children. This means that parents at these centers likely would have improved more had they not been randomly assigned to or attended Head Start. These findings reveal that while Head Start is effective on average at its goal of improving parents’ literacy and math activities with their children, individual Head Start programs vary considerably in their ability to do so. This is an important finding—Head Start has historically been treated in both the academic and policy arenas as a single, uniform program that is operated in many places. The present study provides a deeper understanding of the potential for Head Start to improve families’ outcomes by demonstrating that while Head Start centers operate under a uniform funding structure and set of standards, they should not all be considered uniform in their effectiveness at improving either parent (or child, as previously demonstrated) outcomes.

Notably, Head Start’s impact on the likelihood of parents reading to children daily was consistently positive across programs, suggesting that Head Start is universally good at relaying to parents the importance of daily reading with young children. Perhaps it is easier or more straightforward for providers to convey the importance of daily reading with children than messages about multiple literacy and math

activities. Given the known positive links between regular parent-child reading and children's language and literacy skills (Frijters et al., 2000; Price & Kalil, 2018; Sénéchal & LeFevre, 2002), these are promising findings that suggest that one way that Head Start may affect children's early literacy skills (Puma et al., 2010) is through its impact on parenting.

Daily reading is not the only important cognitively stimulating input to child development, however. Children who are exposed to rich and complex language input and who practice literacy and math concepts and skills regularly tend to develop better language, literacy, and math skills and have higher academic achievement (Crosnoe et al., 2010; Melhuish et al., 2008; Price, 2010; Song et al., 2014; Tamis-LeMonda et al., 2014; Villena-Roldán & Ríos-Aguilar, 2012). Promoting parents' engagement in math activities with children is particularly important because parents are often less confident in their abilities to support early math compared to reading concepts (Levine et al., 2010). It is thus promising that many centers are quite effective at promoting these types of learning-enriching activities between parents and children. Results demonstrate that Head Start nonetheless still has room to grow in these domains such that all Head Start children are supported by their parents in their early learning and development.

Head Start's impacts on parents' physical discipline behaviors were minimal. Only enrollment, and not random assignment, yielded an average treatment effect for parents' likelihood of spanking their children. At first glance, this finding seems to contradict prior studies finding that Head Start reduces parents' use of physical punishment (Puma et al., 2010). However, in the HSIS final impact report, effects of random assignment on spanking were found only for the 3-year-old cohort—these parents were about 7 percentage points less likely to spank their children than control group parents (Puma et al., 2010). No effects on parents' spanking were found in the 4-year-old cohort. Given that the 3-year-old and the 4-year-old cohorts were combined in the present study, it is perhaps not surprising that the already modest average effects on spanking were smaller than in the original impact report and not statistically significant. Results from the present study also indicate that Head Start effects on physical discipline varied only minimally across programs—the only disciplinary treatment effect to vary significantly was the effect of random assignment on parents' frequency of spanking, and the magnitude of this standard deviation was very small. Thus, Head Start was shown to be only minimally effective at reducing physical punishment on average, and while these average effects on spanking may be stronger for 3-year-olds compared to 4-year-olds (Puma et al., 2010), full-sample treatment effects do not vary across centers. Rather, across centers, Head Start treatment effects on parents' spanking use are consistently very small or null,

signaling that most Head Start sites are not effective at changing the overall population of Head Start parents' physical discipline behaviors.

When results from both parenting domains are taken together, they demonstrate that Head Start is effective on average at promoting parents' cognitive stimulation but that these effects still vary widely across centers, while most Head Start centers had little to no success at reducing parents' use of spanking as a disciplinary strategy. That Head Start is better at promoting parents' enriching activities with children than their disciplinary behaviors is perhaps not surprising given that Head Start's main mission is to prepare children for school. It therefore makes sense that Head Start would prioritize the domain of parenting more closely connected to promoting children's early academic skills. Furthermore, discipline as a focus of intervention is often a more sensitive subject than cognitive stimulation for a few reasons. First, the wide held belief in the United States that decisions regarding how to discipline a child should be made privately by parents might lead Head Start centers to decline from providing suggestions for nonphysical discipline, or might make parents less susceptible to these suggestions. Indeed, in a recent report describing interactions between Head Start providers and parents using data from the most recent wave of the Head Start Family and Child Experiences Survey (2014), neither teachers nor family service staff reported talking to parents about discipline, and the majority of both groups reported that they sometimes find it hard to support the way parents choose to discipline their children (Aikens et al., 2017). Second, widespread messages directed at parents about the importance of reading and doing other enriching behaviors with children are more likely to have reached parents across income levels, while there is still a greater reluctance among parents to accept the message that spanking is ineffective and detrimental to children's development—particularly among lower income families (Gershoff et al., 2018). Indeed, despite the documented reduction over time in the overall incidence of spanking for families of all income levels, gaps in spanking use between low- and high-income parents remain such that parents in the 10th income percentile are significantly more likely to spank their children than parents in the 90th income percentile (Ryan et al., 2016). Specifically, in 2011 30% percent of parents in the 10th income percentile endorsed spanking as a disciplinary strategy compared to 12% of parents in the 90th income percentile. Given the link between spanking and negative outcomes like aggression and antisocial behavior for children (Gershoff & Grogan-Kaylor, 2016) as well as the potential for spanking to escalate into more serious abuse (Brown et al., 1998; Sedlak et al., 2010; Zolotor et al., 2008), these levels of spanking use—particularly among lower-income families—are problematic. The American Academy of

Pediatrics has released new guidance for pediatricians and other health care providers for encouraging parents against the use of physical punishment and educating parents about positive and effective alternative disciplinary strategies (Sege et al., 2018). Head Start should consider following these recommendations and begin encouraging providers to educate parents about the harmful effects and ineffectiveness of spanking and to more actively dissuade parents from engaging in spanking as a disciplinary strategy.

While this study makes important contributions to the literature, it is not without its limitations. First, although parenting items and scales like the one used in the current study are commonly used to measure parenting behavior in the home, the study might have been strengthened if observational ratings of parenting quality dimensions such as parental scaffolding, sensitivity, and negative regard were available. Second, while random assignment is the gold standard in evaluation research and allows for stronger causal inferences, the comparison of randomly assigned groups means it is not entirely clear to whom Head Start attendees are being compared because families in the control condition could choose a range of alternative arrangements. Thus, the control condition to which Head Start attendees are compared is made up of children in a number of different arrangements, making it an imprecise comparison. This means that it is not possible to produce causal estimates of Head Start compared to specific ECE arrangement types, though prior work demonstrates that estimates may vary depending on which alternative care arrangement Head Start is being compared to (Feller et al., 2016; Morris et al., 2018; Zhai et al., 2013). For instance, Feller et al. (2016) used the HSIS and found that it was children who would have received care in a home-based setting had they been assigned to the control condition who benefitted most from Head Start (Feller et al., 2016; Morris et al., 2018). By contrast, effects were minimal for children who would have received care in a center-based setting had they been assigned to the control condition. The counterfactual care setting is arguably less important for parenting compared to child outcomes, however, given that the contrast between what Head Start parents and parents in other center-based arrangements receive is likely larger than the contrast between what children in those arrangement types receive since Head Start places a larger emphasis on parent engagement than do other care arrangements. Nonetheless, if the difference between services that Head Start parents receive and services that parents at other arrangements receive is larger for some centers than it is for others, this could affect variation in Head Start's treatment effects on parenting. Future research should therefore consider the possibility that children's counterfactual care condition might also help to explain variation in Head Start treatment effects on parenting, as it does for effects on child outcomes.

Third, while this article makes an important contribution to the literature on Head Start's ability to influence parenting behaviors and to the treatment effect heterogeneity literature, it was unfortunately beyond the scope of the present study to examine predictors of Head Start treatment effect variation on parenting behavior. A natural next question in this line of inquiry is this: What factors predict which centers are more versus less effective? There may be certain characteristics of centers—such as the types of parenting programming that they offer—or of families served at centers—such as the degree to which families face economic risk—that might predict which centers are particularly effective and which are particularly ineffective. Exploring predictors of Head Start treatment effect variation on parenting behavior is a key direction for future work.

Finally, while the HSIS is the ideal data set with which to answer questions about Head Start treatment effect heterogeneity, the data are somewhat old—children in the study attended Head Start in the 2002–2003 school year. This is relevant for two reasons. First, with the wide expansion of public preK programs, the Head Start treatment contrast has changed since 2002. More specifically, the control group in 2002 may have had access to a very different set of options than children today and thus findings may or may not be generalizable to today. Head Start still leads the ECE landscape in terms of its attention to parents, though, so the contrast between outcomes for parents whose children attend Head Start and those who do not today versus in 2002 may not be as different as the contrast in outcomes for children during the same time frame. Second, parenting itself has changed since the time the HSIS was carried out, with parents of all income levels doing significantly more literacy and math activities with children in 2012 than in 1998 (Kalil et al., 2016). The effect of Head Start on parenting behaviors may thus be different today than it was in 2002. Head Start may be less effective at improving parenting if parents served by the program are already doing all they can and do not have as much room to grow, but it may also be more effective at changing parenting if low-income Head Start parents today are more motivated to be involved in their children's learning.

Conclusions

Despite the limitations, this study demonstrates that the previously reported small effects of Head Start on parents' cognitively stimulating interactions with children masked significant variation in the extent to which individual programs modify parents' engagement in literacy and math activities with their children, with some programs much more and some much less effective than their local alternatives. This variation is an important finding worthy of further consideration. Political support and funding for public programs are often determined by the answer to the

question “Does the program work?” The variation demonstrated in the present study contributes to the growing body of evidence (Morris et al., 2018; Weiss et al., 2017) that shows that once a program is taken to scale and implemented in many different locations, one cannot expect to get a uniform answer to that question. Head Start specifically has a central federal funding source but is ultimately operated at the local level, meaning that Head Start funding is distributed from the federal government directly to locally operated grantees. As a result, individual Head Start programs have freedom with regard to the programming they offer, the curricula they use, and the ways in which they choose to engage families, which results in variation in these features across programs (Aikens et al., 2017; U.S. Department of Health and Human Services, Administration for Children and Families, 2016, 2018a). Head Start centers also vary inherently in that they operate in many different locations and thus serve different populations of families. The present study establishes that there is meaningful variation in centers’ effectiveness at promoting positive parenting outcomes, a finding that has been previously established for Head Start’s effects on children’s outcomes (Bloom & Weiland, 2015; Walters, 2015; Weiss et al., 2017). Thus, the answer to the question “Does Head Start work at improving parents’ interactions with their children?” is this: sometimes yes, and sometimes no. As noted above, future research should identify which Head Start program features specifically, or which Head Start parent populations, are most effective at promoting positive parenting behaviors.

Currently, the most common public policy approach to promoting positive parenting behaviors is home visiting (Ryan & Padilla, 2018). Federal funding of home visiting operates through the Maternal, Infant, and Early Childhood Home Visiting (MIECHV) program, which provides funding to states to operate evidence-based home visiting programs (Michalopoulos et al., 2019). The most recent Mother and Infant Home Visiting Program Evaluation (MIHOPE) of four of the most widely used MIECHV programs (Early Head Start—Home-based option, Healthy Families America, Nurse-Family Partnership, and Parents as Teachers) found either no or very small effects of programs on parenting outcomes such as support for learning and literacy, overall supportiveness, and parents’ disciplinary practices (Michalopoulos et al., 2019). Results from the present study thus lend support for the idea that Head Start may be among the most promising avenues for supporting positive parenting practices among low-income families, particularly at the most effective Head Start centers. For example, the MIHOPE evaluation found that home visiting positively impacted parents’ support of literacy and learning, with an average effect size of 0.08 (Michalopoulos et al., 2019). By contrast, I found an average effect of Head Start on parents’ literacy activities of 0.16, with some programs’ effect sizes as large as 0.52. These positive and

relatively large effects combined with both the high rates of ECE use among low-income families with young children, as well as the much larger public investment in Head Start compared to home visiting programs—in 2019 the federal government appropriated \$10 billion for Head Start (U.S. Department of Health and Human Services, Administration for Children and Families, 2019) compared to \$351 million for MIECHV home visiting programs (Health Resources & Services Administration, 2019)—position Head Start at the forefront of promising public approaches to improving parents’ interactions with their young children.

In addition to finding very small average effects of home visiting on parenting outcomes, the MIHOPE evaluation also found that impacts on parenting outcomes did not vary across the four evidence-based programs included in the evaluation (though it is possible that there was variation within programs; this was not included in the evaluation). This lack of variation across programs—particularly for a measure similar to the literacy activities measure included in the present study—is noteworthy given that these were four unique programs operating from distinct (though similar) program models. By contrast, the present study demonstrates variation within a program with a single (though personalizable) program model. The contrast of these two findings further supports the notion that Head Start should by no means be thought of as a uniform program with uniform outcomes. Rather, individual Head Start centers have the potential to be more different from one another than four entirely different programs, all with the same goals of improving parents’ interactions with their children.

As discussed, Head Start is currently prioritizing parent engagement through the PFCE framework, which aims to increase programs’ engagement with parents in order to more fully involve them in their children’s learning and development. However, to the extent that some Head Start programs are quite effective at promoting parenting practices known to enhance children’s school readiness, and others are not effective and even less effective than local alternatives, Head Start decision makers should prioritize understanding what makes some centers more effective than others. We cannot assume that all centers—despite operating from the same funding source and program model—will be equally effective in supporting parents in their children’s development. Rather, individual Head Start centers may need different supports in making this goal a reality. The present study took the first important step by establishing that there is meaningful variation in effects on parents across sites. Future studies should examine which program features, and which parent populations, make some programs more effective than others.

ORCID iD

Christina M. Padilla  <https://orcid.org/0000-0002-7214-2425>

References

- Aikens, N., Cavadel, E., Hartog, J., Hurwitz, F., Knas, E., Schochet, O., Malone, L., & Tarullo, L. (2017). *Building family partnerships: Family engagement findings from the Head Start FACES study*. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. <https://www.acf.hhs.gov/opre/resource/building-family-partnerships-family-engagement-findings-from-the-head-start-faces-study>
- Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*, 444–455. <https://doi.org/10.1080/01621459.1996.10476902>
- Ansari, A., Purtell, K. M., & Gershoff, E. T. (2016). Parenting gains in Head Start as a function of initial parenting skill. *Journal of Marriage and Family*, *78*(5), 1195–1207. <https://doi.org/10.1111/jomf.12296>
- Bandura, A. (1977). Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. <https://doi.org/10.1037/0033-295X.84.2.191>
- Bloom, H. S., Raudenbush, S. W., Weiss, M. J., & Porter, K. (2017). Using multisite experiments to study cross-site variation in treatment effects: A hybrid approach with fixed intercepts and a random treatment coefficient. *Journal of Research on Educational Effectiveness*, *10*(4), 817–842. <https://doi.org/10.1080/19345747.2016.1264518>
- Bloom, H. S., & Weiland, C. (2015). *Quantifying variation in Head Start effects on young children's cognitive and socio-emotional skills using data from the National Head Start Impact Study*. MDRC.
- Brachet, T. (2007). *Documentation for computing clustered standard errors for two-stage least squares in SAS*. <http://works.bepress.com/tbrachet/2/>
- Bradley, R. H., Corwyn, R. F., Burchinal, M., McAdoo, H. P., & Garcia Coll, C. (2001). The home environments of children in the United States part II: Relations with behavioral development through age thirteen. *Child Development*, *72*(6), 1868–1886. <https://doi.org/10.1111/1467-8624.t01-1-00383>
- Brooks-Gunn, J., Berlin, L. J., & Fuligni, A. (2000). Early childhood intervention programs: What about the family? In S. Meisels, & J. Shonkoff (Eds.), *The handbook of early intervention* (2nd ed., pp. 549–588). Cambridge University Press.
- Brown, J., Cohen, P., Johnson, J. G., & Salzinger, S. (1998). A longitudinal analysis of risk factors for child maltreatment: findings of a 17-year prospective study of officially recorded and self-reported child abuse and neglect. *Child Abuse & Neglect*, *22*(11), 1065–1078. [https://doi.org/10.1016/S0145-2134\(98\)00087-8](https://doi.org/10.1016/S0145-2134(98)00087-8)
- Cabrera, N. J., Ryan, R. M., Shannon, J., Brooks-Gunn, J., Vogel, C., Raikes, S., Tamis-LeMonda, C., & Cohen, R. (2004). Low-income fathers' involvement in their toddlers' lives: Biological fathers from the Early Head Start research and evaluation study. *Fathering: A Journal of Theory, Research, and Practice About Men as Fathers*, *2*(1), 5–36. <https://doi.org/10.3149/ftf.0201.5>
- Cain, G. G. (1975). Regression and selection models to improve nonexperimental comparisons. In C. A. Bernet, & A. A. Lumsdiane (Eds.), *Evaluation and experiment* (pp. 297–317). Academic Press.
- Caldwell, B. M., & Bradley, R. H. (2003). *Administration manual: Home observation for measurement of the environment*. University of Arkansas at Little Rock.
- Chase-Lansdale, P. L., Moffitt, R. A., Lohman, B. J., Cherlin, A. J., Coley, R. L., Pittman, L. D., Roff, J., & Votruba-Drzal, E. (2003). Mothers' transitions from welfare to work and the well-being of preschoolers and adolescents. *Science*, *299*(5612), 1548–1552. <https://doi.org/10.1126/science.1076921>
- Chase-Lansdale, P. L., Sabol, T. J., Sommer, T. E., Chor, E., Cooperman, A. W., Brooks-Gunn, J., Yoshikawa, H., King, C., & Morris, A. (2019). Effects of a two-generation human capital program on low-income parents' education, employment, and psychological wellbeing. *Journal of Family Psychology*, *33*(4), 433–443. <https://doi.org/10.1037/fam0000517>
- Chaudry, A., Pedroza, J., & Sandstrom, H. (2012). *Perspectives on low-income working families: How employment constraints affect low-income working parents' child care decisions* (Brief No. 23). Urban Institute.
- Conger, R. D., & Conger, K. J. (2002). Resilience in midwestern families: Selected findings from the first decade of a prospective, longitudinal study. *Journal of Marriage and Family*, *64*(2), 361–373. <https://doi.org/10.1111/j.1741-3737.2002.00361.x>
- Crosnoe, R., Leventhal, T., Wirth, R. J., Pierce, K. M., & Pianta, R. C. (2010). Family socioeconomic status and consistent environmental stimulation in early childhood. *Child Development*, *81*(3), 972–987. <https://doi.org/10.1111/j.1467-8624.2010.01446.x>
- Dodge, K. A., Pettit, G. S., & Bates, J. E. (1994). Socialization mediators of the relation between socioeconomic status and child conduct problems. *Child Development*, *65*(2), 649–665. <https://doi.org/10.1111/j.1467-8624.1994.tb00774.x>
- Enchautegui, M. E. (2013). *Nonstandard work schedules and the well-being of low-income families* (Paper 26). Urban Institute.
- Feller, A., Grindal, T., Miratrix, L., & Page, L. C. (2016). Compared to what? Variation in the impacts of early childhood education by alternative care type. *Annals of Applied Statistics*, *10*(3), 1245–1285. <https://doi.org/10.1214/16-AOAS910>
- Frijters, J. C., Barron, R. W., & Brunello, M. (2000). Direct and mediated influences of home literacy and literacy interest on prereaders' oral vocabulary and early written language skill. *Journal of Educational Psychology*, *92*(3), 466–477. <https://doi.org/10.1037/0022-0663.92.3.466>
- Gelber, A., & Isen, A. (2013). Children's schooling and parents' behavior: Evidence from the Head Start Impact Study. *Journal of Public Economics*, *101*, 25–38. <https://doi.org/10.1016/j.jpubeco.2013.02.005>
- Gershoff, E. T., Goodman, G. S., Miller-Perrin, C. L., Holden, G. W., & Kazdin, A. E. (2018). The strength of the causal evidence against physical punishment of children and its implications for parents, psychologists, and policymakers. *American Psychologist*, *73*(5), 626–638. <https://doi.org/10.1037/amp0000327>
- Gershoff, E. T., & Grogan-Kaylor, A. (2016). Spanking and child outcomes: Old controversies and new meta-analyses. *Journal of Family Psychology*, *30*(4), 453–469. <https://doi.org/10.1037/fam0000191>
- Gershoff, E. T., Grogan-Kaylor, A., Lansford, J. E., Chang, L., Zelli, A., Deater-Deckard, K., & Dodge, K. A. (2010). Parent discipline practices in an international sample: Associations

- with child behaviors and moderation by perceived normativeness. *Child Development*, 81(2), 487–502. <https://doi.org/10.1111/j.1467-8624.2009.01409.x>
- Guryan, J., Hurst, E., & Kearney, M. S. (2008). Parent education and parental time with children. *Journal of Economic Perspectives*, 22(2), 23–46. <https://doi.org/10.1257/jep.22.3.23>
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experiences of young American children*. Brookes.
- Health Resources & Services Administration. (2019). *Maternal, Infant, and Early Childhood Home Visiting awards FY19*. <https://mchb.hrsa.gov/maternal-child-health-initiatives/home-visiting/fy19-awards>
- Heath, S. M., Wigley, C. A., Hogben, J. H., Fletcher, J., Collins, P., Boyle, G. L., & Eustice, S. (2018). Patterns in participation: Factors influencing parent attendance at two, centre-based early childhood interventions. *Journal of Child and Family Studies*, 27(1), 253–267. <https://doi.org/10.1007/s10826-017-0878-2>
- Heckman, J. J., & Robb, R. (1985). Alternative methods for evaluating the impact of an intervention. *Journal of Econometrics*, 30, 239–267. [https://doi.org/10.1016/0304-4076\(85\)90139-3](https://doi.org/10.1016/0304-4076(85)90139-3)
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta analysis*. Academic Press.
- Howard, K. S., & Brooks-Gunn, J. (2009). The role of home-visiting programs in preventing child abuse and neglect. *The Future of Children*, 19(2), 119–146. <https://doi.org/10.1353/foc.0.0032>
- Improving Head Start for School Readiness Act of 2007. (2007). 110th Congress. <https://eclkc.ohs.acf.hhs.gov/sites/default/files/pdf/hs-act-pl-110-134.pdf>
- Kalil, A., & Ryan, R. (2020). Parenting practices and socioeconomic gaps in childhood outcomes. *The Future of Children*, 30(1), 29–54. <https://doi.org/10.1353/foc.2020.0004>
- Kalil, A., Ryan, R. M., & Corey, M. (2012). Diverging destinies: Maternal education and the developmental gradient in time with children. *Demography*, 49(4), 1361–1383. <https://doi.org/10.1007/s13524-012-0129-5>
- Kalil, A., Ziol-Guest, K. M., Ryan, R. M., & Markowitz, A. J. (2016). Changes in income-based gaps in parent activities with young children from 1988 to 2012. *AERA Open*, 2(3), 1–17. <https://doi.org/10.1177/2332858416653732>
- Kessler, R., & Greenberg, E. F. (1981). *Linear panel analysis: Models of quantitative change*. Academic Press.
- Leavell, A. S., Tamis-LeMonda, C. S., Ruble, D. N., Zosuls, K. M., & Cabrera, N. J. (2012). African American, White and Latino fathers' activities with their sons and daughters in early childhood. *Sex Roles*, 66(1–2), 53–65. <https://doi.org/10.1007/s11199-011-0080-8>
- Lee, R., Brooks-Gunn, J., Han, W.-J., Waldfogel, J., & Zhai, F. (2014). Is participation in Head Start associated with less maternal spanking for boys and girls? *Children and Youth Services Review*, 46, 55–63. <https://doi.org/10.1016/j.childyouth.2014.08.006>
- Levine, S. C., Suriyakham, L. W., Rowe, M. L., Huttenlocher, J., & Gunderson, E. A. (2010). What counts in the development of young children's number knowledge? *Developmental Psychology*, 46(5), 1309–1319. <https://doi.org/10.1037/a0019671>
- Magnuson, K. A., & Waldfogel, J. (2005). Preschool child care and parents' use of physical discipline. *Infant and Child Development*, 14(2), 177–198. <https://doi.org/10.1002/icd>
- Mayer, S. E., Kalil, A., Oreopoulos, P., & Gallegos, S. (2015). *Using behavioral insights to increase parental engagement: The Parents and Children Together (PACT) intervention* (NBER Working Paper Series, Working Paper 21602). National Bureau of Economic Research. <http://www.nber.org/papers/w21602>
- McCartney, K., Dearing, E., Taylor, B. A., & Bub, K. L. (2007). Quality child care supports the achievement of low-income children: Direct and indirect pathways through caregiving and the home environment. *Journal of Applied Developmental Psychology*, 28(5–6), 411–426. <https://doi.org/10.1016/j.appdev.2007.06.010>
- McCoy, D. C., Morris, P. A., Connors, M. C., Gomez, C. J., & Yoshikawa, H. (2016). Differential effectiveness of Head Start in urban and rural communities. *Journal of Applied Developmental Psychology*, 43, 29–42. <https://doi.org/10.1016/j.appdev.2015.12.007>
- Melhuish, E. C., Phan, M. B., Sylva, K., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2008). Effects of the home learning environment and preschool center experience upon literacy and numeracy development in early primary school. *Journal of Social Issues*, 64(1), 95–114. <https://doi.org/10.1111/j.1540-4560.2008.00550.x>
- Michalopoulos, C., Faucetta, K., Hill, C. J., Portilla, X. A., Burrell, L., Lee, H., Duggan, A., & Knox, V. (2019). *Impacts on family outcomes of evidence-based early childhood home visiting: Results from the Mother and Infant Home Visiting Program evaluation* (OPRE Report 2019-07). Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Mirowsky, J., & Ross, C. E. (2003). *Education, social status and health*. Aldine de Gruyter.
- Moiduddin, E., Aikens, N., Tarullo, L., West, J., & Xue, Y. (2012). *Child outcomes and classroom quality in FACES* (OPRE Report 2012-37a). Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Morris, P. A., Connors, M., Friedman-krauss, A., McCoy, D. C., Weiland, C., & Bloom, H. (2018). New findings on impact variation from the Head Start Impact Study: Informing the scale-up of early childhood programs. *AERA Open*, 4(2), 1–16. <https://doi.org/10.1177/2332858418769287>
- Mytton, J., Ingram, J., Manns, S., Hons, L., & Thomas, J. (2014). Facilitators and barriers to engagement in parenting programs: A qualitative systematic review. *Health Education & Behavior*, 41(2), 127–137. <https://doi.org/10.1177/1090198113485755>
- National Institute of Child Health and Human Development Early Child Care Research Network. (2004). Fathers' and mothers' parenting behavior and beliefs as predictors of children's social adjustment in the transition to school. *Journal of Family Psychology*, 18(4), 628–638. <https://doi.org/10.1037/0893-3200.18.4.628>
- National Institute of Child Health and Human Development Early Child Care Research Network & Duncan, G. J. (2003). Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development*, 74(5), 1454–1475. <https://doi.org/10.1111/1467-8624.00617>
- Office of Head Start, Administration for Children and Families. (2019). *Head Start programs*. <https://www.acf.hhs.gov/ohs/about/head-start>

- Padilla, C. M., & Ryan, R. M. (2019). The link between child temperament and low-income mothers' and fathers' parenting. *Infant Mental Health Journal, 40*(2), 217–233. <https://doi.org/10.1002/imhj.21770>
- Phillips, M. (2011). Parenting, time use, and disparities in academic outcomes. In G. Duncan & R. Murnane (Eds.), *Whither opportunity: Rising inequality, schools, and children's life chances* (pp. 207–228). Russell Sage and Spencer Foundation.
- Price, J. (2010). The effects of parental time investments: Evidence from natural within-family variation (Brigham Young University working paper). http://byuresearch.org/home/downloads/price_parental_time_2010.pdf
- Price, J., & Kalil, A. (2018). The effect of mother-child reading time on children's reading skills: Evidence from natural within-family variation. *Child Development, 90*(6), e688–e702. <https://doi.org/10.1111/cdev.13137>
- Prinz, R. J., & Sanders, M. R. (2007). Adopting a population-level approach to parenting and family support interventions. *Clinical Psychology Review, 27*(6), 739–749. <https://doi.org/10.1016/j.cpr.2007.01.005>
- Puma, M., Bell, S., Cook, R., & Heid, C. (2010). *Head Start Impact Study final report*. Administration for Children and Families, Office of Planning, Research and Evaluation.
- Rafferty, Y., Griffin, K. W., & Robokos, D. (2010). Maternal depression and parental distress among families in the Early Head Start Research and Evaluation Project: Risk factors within the family setting. *Infant Mental Health Journal, 31*(5), 543–569. <https://doi.org/10.1002/imhj.20271>
- Raudenbush, S. W., & Bloom, H. S. (2015). Learning about and from a distribution of program impacts using multisite trials. *American Journal of Evaluation, 36*(4), 475–499. <https://doi.org/10.1177/1098214015600515>
- Raudenbush, S. W., Reardon, S. F., & Nomi, T. (2012). Statistical analysis for multisite trials using instrumental variables with random coefficients. *Journal of Research on Educational Effectiveness, 5*(3), 303–332. <https://doi.org/10.1080/19345747.2012.689610>
- Rowe, M. L. (2018). Understanding socioeconomic differences in parents' speech to children. *Child Development Perspectives, 12*(2), 122–127. <https://doi.org/10.1111/cdep.12271>
- Royston, P. (2005). Multiple imputation of missing values: Update of ice. *Stata Journal, 5*(4), 527–536. <https://doi.org/10.1177/1536867X0500500404>
- Ryan, R. M., Kalil, A., Ziol-Guest, K. M., & Padilla, C. (2016). Socioeconomic gaps in parents discipline strategies from 1988 to 2011. *Pediatrics, 138*(6), e20160720–e20160720. <https://doi.org/10.1542/peds.2016-0720>
- Ryan, R. M., & Padilla, C. M. (2018). Public policy and family psychology. In B. H. Fiese (Ed.), *APA handbook of contemporary family psychology*. American Psychological Association.
- Sabol, T. J., & Chase-Lansdale, P. L. (2015). The influence of low-income children's participation in Head Start on their parents' education and employment. *Journal of Policy Analysis and Management, 34*(1), 136–161. <https://doi.org/10.1002/pam>
- Sabol, T. J., McCoy, D. C., Gonzalez, K., Miratrix, L., Hedges, L., & Spybrook, J. K. (2020). *For whom, where, and when? Challenges and opportunities in exploring cross-site treatment impact variation and why psychologists should care*. Manuscript submitted for publication.
- Sedlak, A. J., Mettenberg, J., Basena, M., Petta, I., McPherson, K., Greene, A., & Li, S. (2010). *Fourth National Incidence Study of child abuse and neglect* (NIS-4). <https://www.acf.hhs.gov/opre/research/project/national-incidence-study-of-child-abuse-and-neglect-nis-4-2004-2009>
- Sege, R. D., & Siegel, B. S., & Council on Child Abuse and Neglect Committee on Psychosocial Aspects of Child and Family. (2018). Effective discipline to raise healthy children. *Pediatrics, 142*(6), e20183112. <https://doi.org/10.1542/peds.2018-3112>
- Sénéchal, M., & LeFevre, J. A. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development, 73*(2), 445–460. <https://doi.org/10.1111/1467-8624.00417>
- Song, L., Spier, E. T., & Tamis-LeMonda, C. S. (2014). Reciprocal influences between maternal language and children's language and cognitive development in low-income families. *Journal of Child Language, 41*(02), 305–326. <https://doi.org/10.1017/S0305000912000700>
- Suizzo, M., & Stapleton, L. M. (2007). Home-based parental involvement in young children's education: Examining the effects of maternal education across U.S. ethnic groups. *Educational Psychology, 27*(4), 533–556. <https://doi.org/10.1080/01443410601159936>
- Tamis-LeMonda, C. S., Song, L., Luo, R., Kuchirko, Y., Kahana-Kalman, R., Yoshikawa, H., & Raufman, J. (2014). Children's vocabulary growth in English and Spanish across early development and associations with school readiness skills. *Developmental Neuropsychology, 39*(2), 69–87. <https://doi.org/10.1080/87565641.2013.827198>
- U.S. Department of Health and Human Services, Administration for Children and Families. (n.d.). *Office of Head Start fact sheet*. https://www.acf.hhs.gov/sites/default/files/ohs/factsheets_ohs.pdf
- U.S. Department of Health and Human Services, Administration for Children and Families. (2016). *Head Start program performance standards*. <https://eclkc.ohs.acf.hhs.gov/sites/default/files/pdf/hspss-final.pdf>
- U.S. Department of Health and Human Services, Administration for Children and Families. (2018a). *Head Start Parent, Family, and Community Engagement Framework*. <https://eclkc.ohs.acf.hhs.gov/school-readiness/article/head-start-parent-family-community-engagement-framework>
- U.S. Department of Health and Human Services, Administration for Children and Families. (2018b). *Head Start Program facts: Fiscal Year 2018*. <https://eclkc.ohs.acf.hhs.gov/about-us/article/head-start-program-facts-fiscal-year-2018>
- U.S. Department of Health and Human Services, Administration for Children and Families. (2019). *Head Start Program facts Fiscal Year 2019*. <https://eclkc.ohs.acf.hhs.gov/about-us/article/head-start-program-facts-fiscal-year-2019>
- Villena-Roldán, B., & Ríos-Aguilar, C. (2012). *Causal effects of maternal time-investment on children's cognitive outcomes* (Working Paper 285). Center for Applied Economics.
- Votruba-Drzal, E., & Chase-Lansdale, P. L. (2004). Child care and low-income children's development: Direct and moderated effects. *Child Development, 75*(1), 296–312. <https://doi.org/10.1111/j.1467-8624.2004.00670.x>
- Waldfoegel, J., & Washbrook, E. (2011). Early years policy. *Child Development Research, 2011*(1), Article ID 343016. <https://doi.org/10.1155/2011/343016>

- Walters, C. R. (2015). Inputs in the production of early childhood human capital: Evidence from Head Start. *American Economic Journal*, 7(4), 76–102. <https://doi.org/10.1257/app.20140184>
- Weiss, M. J., Bloom, H. S., & Brock, T. (2014). A conceptual framework for studying the sources of variation in program effects. *Journal of Policy Analysis and Management*, 33(3), 778–808. <https://doi.org/10.1002/pam.21760>
- Weiss, M. J., Bloom, H. S., Verbitsky-Savitz, N., Gupta, H., Vigil, A. E., & Cullinan, D. N. (2017). How much do the effects of education and training programs vary across sites? Evidence from past multisite randomized trials. *Journal of Research on Educational Effectiveness*, 10(4), 843–876. <https://doi.org/10.1080/19345747.2017.1300719>
- York, B. N., & Loeb, S. (2014). *One step at a time: The effects of an early literacy text messaging program for parents of pre-schoolers* (NBER Working Paper Series, Working Paper No. 20659). National Bureau of Economic Research.
- Zhai, F., Waldfogel, J., & Brooks-Gunn, J. (2013). Estimating the effects of Head Start on parenting and child maltreatment. *Children and Youth Services Review*, 35(7), 1119–1129. <https://doi.org/10.1016/j.chilyouth.2011.03.008>
- Zigler, E., & Styfco, S. J. (2004). *The Head Start debates*. Brookes.
- Zigler, E., & Valentine, J. (1979). *Project Head Start: A legacy of the war on poverty*. Free Press. <https://eric.ed.gov/?id=ED183266>
- Zolotor, A. J., Theodore, A. D., Chang, J. J., Berkoff, M. C., & Runyan, D. K. (2008). Speak softly — and forget the stick: Corporal punishment and child physical abuse. *American Journal of Preventive Medicine*, 35(4), 364–369. <https://doi.org/10.1016/j.amepre.2008.06.031>

Author

CHRISTINA M. PADILLA completed her doctoral studies at Georgetown University and is currently a research scientist at Child Trends. She is interested in the independent and joint influences of parenting and early childhood education on children's early development.