MOVING BEYOND THE AVERAGE EFFECT:
QUANTIFYING AND EXPLORING VARIATION IN HEAD START
TREATMENT EFFECTS ON PARENTING BEHAVIOR

A Dissertation
Submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Psychology

By

Christina Michelle Louise Padilla, M.P.P.

Washington, D.C.
April 5, 2019
MOVING BEYOND THE AVERAGE EFFECT: QUANTIFYING AND EXPLORING VARIATION IN HEAD START TREATMENT EFFECTS ON PARENTING BEHAVIOR

Christina M. Padilla, M.P.P.

Thesis Advisor: Rebecca M. Ryan, Ph.D.

ABSTRACT

Differences in the quality and quantity of parenting that children in low- compared to middle- and high-socioeconomic status (SES) families experience account for a quarter to a third of the gaps between these groups of children in their early skills (Waldfogel & Washbrook, 2011). Parenting interventions hold promise for promoting positive parenting practices among low-income parents, but face barriers with regard to recruitment and retention. As one of the largest sources of federal support for low-income families in the U.S., Head Start provides a promising opportunity to enhance parenting in this population. Prior research has found evidence for small-moderate impacts of Head Start on some parenting behaviors (Gelber & Isen, 2013; Puma, Bell, Cook, & Heid, 2010). However, prior research focuses on the average effect of Head Start, ignoring potential heterogeneity of treatment effects across different Head Start centers. In a set of three studies, the present dissertation quantifies the cross-site variation in Head Start treatment effects on parenting behaviors and tests whether that variation is statistically different from zero; explores whether characteristics of centers predict variation in treatment effects; and explores whether characteristics of families served by centers predict variation in treatment effects. Findings demonstrated that Head Start is effective on average at promoting parents’ cognitively stimulating behaviors with children but that programs vary in their effectiveness, and that Head Start is consistently ineffective at lowering parents’ use of harsh punishment. Some of the examined center and family characteristics predict cross-site variation in Head Start effects, but sometimes in unexpected directions. Implications of these findings are discussed.
ACKNOWLEDGEMENTS

They say it takes a village to raise a child, and it certainly took a village to make this dissertation possible.

First, I would like to thank all of the families, children, teachers, and Head Start providers who participated in the Head Start Impact Study, as well as all those who worked to collect, clean, and share the data. This dissertation would quite literally not be possible without them.

Thank you to the Doris Duke Fellowships for the Promotion of Child Well-Being for not only providing me with two years of generous dissertation support, but also for exposing me to a brilliant network of scholars, who will be my colleagues for the rest of my career.

Thank you to those who have lent their time and expertise along the way. Dr. Terri Sabol, thank you for agreeing to join my committee, for your time and encouragement, for sharing your knowledge, and most importantly, for reminding me that this work is important. Katie Gonzalez, Dr. Mike Weiss, and Dr. Christina Weiland, thank you for your time and methodological expertise.

I would like to thank the members of the Georgetown Psychology Department for providing an environment in which I could learn about both neurons and neighborhoods, and everything in between.

To my parents, thank you for believing in me always and for inspiring and fostering my love of learning from the beginning. Your sacrifices have made all of this possible, and I am forever grateful.

To my friends, I cannot thank you enough for the incredible amount of support you have given me these past six years. Emily, Holly, and Sam, you three were such an unexpected but treasured gift from my time at McCourt. You have kept up with the ins and outs of this dissertation coming together, and you have kept me afloat with an endless amount of encouragement and support. You are the only people to inquire about seemingly mundane issues such as the progress I’ve made on the HSIS technical appendix, and I cannot thank you enough. Circle power forever! Katie, you have been my life coach since we met in 2011. I am so lucky to have found both friendship and mentorship in you—thank you for both. To my Hopkins crew, thank you for being my happy place always.

To my friends and colleagues in the Psychology Department, you have been invaluable during this journey. Anna Markowitz, I learned so much from your example. You showed me how to be a passionate and dedicated thinker while always keeping the mission at the forefront of our work. Thank you for your friendship and support along the way. Elise, you have been the most wonderful cheerleader and reminder that I could and would finish this thing. Thank you for teaching me about work-life balance, for lending me your cat to snuggle, and for the best hugs in the world. Marisa and Kruti, I cannot thank you enough for your constant willingness to listen (often during much needed walks home), for providing a place to bounce around ideas, and for your companionship during this crazy time in life.
I have been incredibly fortunate to have been part of such a brilliant and collegial lab. Caitlin, you are a truly wonderful colleague and friend. Thank you for making my life easier and better (in work and in life). Jane, your presence in the lab has brought me so much joy, and I am so grateful that you are my friend. Thank you for making it fun to come in every day. Owen, your eagerness and curiosity is inspiring, and I have learned so much from you. Annie, this past year you have unknowingly reminded me to be excited about our work at times when it was easy to forget—thank you. I can’t wait to see the great things that all of you accomplish and to cheer you on every step of the way.

I owe an immense amount of gratitude to the Child Development and Social Policy lab’s co-directors: Dr. Deborah Phillips, Dr. Anna Johnson, and Dr. Rebecca Ryan. Deborah, your love of our work and your passion for the mission is truly infectious. Thank you for leading by example, for your words of wisdom, for your positivity, and for your encouragement.

Anna, I had no idea when I came to Georgetown that I would come to be mentored by not just one, but two exceptional scholars. Your unexpected contribution to my development as a thinker and as a human means more to me than you will ever know. You have shown me how to be kind, generous, and patient, and your unwavering belief in my abilities has helped me to believe in myself. Thank you for everything.

Rebecca, I cannot put into words what your mentorship has meant to me. You are a brilliant thinker and an encouraging, understanding, and inspiring mentor. You have pushed me to be my best, taught me to be a critical thinker, and to never give up (even after many rejections!). I am so incredibly lucky to have learned from you. Thank you for six years of close and careful mentorship, for believing in me before I had any clue what I was doing, for so many draft reads and re-reads, and for sharing your love of statistics with me. I am so grateful to and for you. Thank you.
# TABLE OF CONTENTS

CHAPTER I: Introduction ........................................................................................................... 1

CHAPTER II: Variation in Head Start program effects on parents’ interactions with children ................................................................................................................................. 24

CHAPTER III: Center characteristics as sources of variation in Head Start effects on parenting outcomes .............................................................................................................................................. 57

CHAPTER IV: Family characteristics as sources of variation in Head Start effects on parenting outcomes .............................................................................................................................................. 97

CHAPTER V: General discussion ................................................................................................................. 130

APPENDIX .............................................................................................................................................. 142

SUPPLEMENTARY TABLES ...................................................................................................................... 144

REFERENCES .......................................................................................................................................... 155
LIST OF FIGURES

Figure 1: Cross-site distributions of Head Start ITT effects by outcome measure............... 55

Figure 2: Cross-site distributions of Head Start ITT effects by outcome measure based on the OLS with treatment-by-center interactions approach (dashed line) and the fixed intercept, random slopes approach (solid line). ................................................................. 56

Figure 3: Variation across centers in outreach activities. .................................................... 92

Figure 4: Variation across centers in quality scores. .......................................................... 93

Figure 5: ITT effects of random assignment on parenting outcomes for continuous outcomes (left) and binary outcomes (right), by Family Service Worker (FSW) Outreach. ............. 94

Figure 6: ITT effects of random assignment on parenting outcomes for continuous outcomes (left) and binary outcomes (right), by number of home visits. .......................................... 95

Figure 7: ITT effects of random assignment on parenting outcomes for continuous outcomes (left) and binary outcomes (right), by Arnett CIS scores..................................................... 96

Figure 8: Variation across centers in household economic risk, maternal immigrant status, multigenerational Head Start status, and parents’ initial cognitive stimulation and spanking. ................................................................................................. 126

Figure 9: ITT effect of random assignment on parenting outcomes for continuous outcomes (left) and binary outcomes (right), by percentage of families within a center with an immigrant mother. ........................................................................................................ 127

Figure 10: ITT effect of random assignment on parenting outcomes for continuous outcomes (left) and binary outcomes (right), by percentage multigenerational Head Start (MGHS) families within a center................................................................. 128
Figure 11: ITT effect of random assignment on parenting outcomes for continuous outcomes (left) and binary outcomes (right), by pre-academic stimulation scores within a center.
LIST OF TABLES

Table 1: Descriptive statistics and baseline balance of the analytic sample .................................. 51
Table 2: Average effects of Head Start random assignment and enrollment on parenting outcomes and standard deviations of effects ................................................................. 51
Table 3: Descriptive statistics of center characteristics ................................................................. 88
Table 4: Pairwise correlations between centers' parent outreach and classroom quality measures .................................................................................................................. 89
Table 5: Models showing the non-causal ITT effect of Head Start parent outreach, home visiting, and classroom quality on the average Head Start effect ................................................. 90
Table 6: Descriptive statistics of center-level family characteristics ............................................. 122
Table 7: Pairwise correlations between center-level family characteristics ................................. 123
Table 8: Models showing the non-causal ITT effect of Head Start centers' average economic risk status, immigrant status, multigenerational Head Start status, and initial parenting on the average Head Start effect ........................................................................................................ 124
Table A1: Models showing the non-causal ITT effect of ECERS-R subscales on the average Head Start effect .................................................................................................................. 144
Table A2: Models showing the non-causal ITT effect of ECERS-R total scores on the average Head Start effect when "high" ECERS-R quality is defined as a score of six or above ................................................................................................................................. 146
Table A3: Descriptive statistics showing child and family level center-level demographic characteristics stratified by high versus low family service worker outreach ........................................................................................................ 147
Table A4: Descriptive statistics showing center-level family demographic characteristics for centers differing in the percentage of moderate-high risk families they serve ............ 148

Table A5: Descriptive statistics showing center-level family demographic characteristics for centers differing in the percentage of immigrant families they serve ......................... 149

Table A6: Descriptive statistics showing center-level family demographic characteristics for centers differing in the percentage of multigenerational Head Start (MGHS) families they serve .................................................................................................................. 151

Table A7: Descriptive statistics showing center-level family demographic characteristics for centers differing in the average initial cognitive stimulation among families they serve .............................................................................................................................................................. 152

Table A8: Descriptive statistics showing center-level family demographic characteristics for centers differing in the percentage of parents who reported initially spanking they serve ........................................................................................................................................................................... 153
CHAPTER 1: INTRODUCTION

Children’s early experiences with their caregivers are one of the most potent influences on their development. Children who frequently experience warm, consistent, and stimulating interactions with their parents tend to have stronger early math, language, and reading skills; better attention regulation and communication skills; and a greater interest in learning than those who lack them (Crosnoe, Leventhal, Wirth, Pierce, & Pianta, 2010; Melhuish et al., 2008; Price, 2010; Song, Spier, & Tamis-LeMonda, 2014; Tamis-LeMonda et al., 2014; Villena-Roldán & Ríos-Aguilar, 2012), while children who experience more frequent negative or intrusive parenting or discipline practices like hitting or spanking tend to experience more negative outcomes than their peers, including aggression, externalizing, delinquent and antisocial behavior, poorer mental health, and impairments in cognitive ability (Gershoff & Grogan-Kaylor, 2016). On average, economically disadvantaged parents parent their children differently than more advantaged parents in ways that have significant implications for their children’s development. For instance, economically advantaged parents tend to provide their children with more language and cognitive stimulation (e.g. Bradley, Burchinal, & Coll, 2001; Hart & Risley, 1995; Kalil, Ryan, & Corey, 2012; Phillips, 2011), tend to behave more sensitively and contingently during mother-child interactions (Conger & Conger, 2002; Dodge, Pettit, & Bates, 1994; NICHD ECCRN, 2004), and tend to spend more time in active care of children (Guryan, Hurst, & Kearney, 2008; Sayer, Bianchi, & Robinson, 2004)—time that is more developmentally appropriate (Kalil et al., 2012) than less advantaged parents. More advantaged parents are also less likely than less advantaged parents to spank children or use harsh discipline (Gershoff & Grogan-Kaylor, 2016; Zolotor, Theodore, Chang, Berkoff, & Runyan, 2008), practices shown to increase the risk of more serious abuse (Brown, Cohen, Johnson, & Salzinger, 1998; Sedlak et
al., 2010). These differences in the quality and quantity of parenting that children in low-compared to middle- and high-socioeconomic status (SES) families experience account for a quarter to a third of the gaps between these groups of children in early literacy, mathematics, and language skills (Waldfogel & Washbrook, 2011). Understanding how to promote the types of parenting behaviors known to boost children’s early academic and behavioral skill development among low-income families is thus important for closing these skill gaps.

Many interventions with the goal of closing or minimizing these gaps attempt to intervene with low-income or otherwise at-risk parents with the purpose of promoting parenting behaviors known to enhance child development, and ultimately, improving child outcomes. These programs—including home visiting and other parenting interventions—can be effective at changing parents’ behaviors (see Ryan & Padilla, 2018 for a review). However, program providers and administrators often report substantial difficulty recruiting and then retaining low-income families in these programs (Heath et al., 2018), often because low-income parents have multiple jobs and irregular employment and child care schedules, making it difficult to attend programs regularly (Chaudry, Pedroza, & Sandstrom, 2012; Enchautegui, 2013; Mytton, Ingram, Manns, Hons, & Thomas, 2014; Prinz & Sanders, 2007). Thus, it is judicious to consider additional ways to reach these families.

Sixty-one percent of families in the United States used some form of child care or early care and education (ECE) in 2011, the most recent year for which Census data are available (Laughlin, 2013). Included in this majority are millions of low-income families, who often relying on publicly funded care such as Head Start, child care subsidies, or public pre-k. Because many low-income families already use ECE, reaching them through ECE services could be more effective than attempts to engage them in additional services. As one of the largest sources of
federal support for low-income families, Head Start provides a promising opportunity to promote positive parenting practices in this population. Head Start was created in 1965 as part of the War on Poverty and is intended to boost the school readiness of children from low-income families (Puma, Bell, Cook, & Heid, 2010). 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). Based on a belief that boosting children’s skills requires attention to multiple aspects of children’s development, Head Start is also based on a “whole child” model and so provides not only preschool education, but also medical, dental, mental health, and nutrition services, as well as attempts 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). For example, the newly updated Head Start performance standards emphasize efforts to involve parents and families in their children’s early learning through its Parent, Family, and Community Engagement Framework (Administration for Children and Families, 2018). This framework emphasizes the shared responsibility of center staff and family members in promoting the healthy development of Head Start children.

Prior studies examining whether Head Start is effective at enhancing parents’ interactions with their children have found evidence for small-moderate impacts on the frequency with which parents engage in reading and math activities with children and their likelihood of spanking children (Ansari, Purtell, & Gershoff, 2016; Gelber & Isen, 2013; Puma et al., 2010). These studies, however, have not considered the likely possibility that individual Head Start programs vary in their effectiveness at improving these and other parenting outcomes. Prior research has
thus not been able to fully elucidate the conditions under which or the families for whom Head Start is most effective at reaching its goal of engaging parents in their children’s development.

This dissertation explores the causal effect of Head Start on parents’ parenting behaviors in two domains: cognitive stimulation and disciplinary behaviors. More specifically, I use a large multi-site randomized-controlled trial (RCT) with rigorous methodological techniques to first replicate prior reports of the average effect of one year of Head Start on changes in parenting, and then to quantify variation in Head Start effects on parents across sites—which no other study has done. I then explore two potential sources of variation in program effects: characteristics of centers themselves, and characteristics of families that centers serve. Results from this dissertation will add to our understanding of the full potential for Head Start to enhance parenting practices, and thus have important policy and practice implications. For instance, results could help Head Start decision-makers to understand where and how efforts to strengthen parenting outcomes might be best directed by identifying those aspects of the program most strongly associated with improvements in parenting. Proliferating those practices throughout Head Start centers and providers could help ensure the program offers the greatest benefits for low-income children and families.

**Parenting Behaviors Targeted by Head Start**

Head Start targets two types of parenting behaviors that have strong theoretical and empirical links with children’s outcomes: cognitively stimulating and socioemotional interactions. Cognitive stimulation includes enriching behaviors like reading, other literacy activities, and discussing math concepts. Positive socioemotional interactions involve parental warmth and consistency as well as a lack of harsh discipline or physical punishment. Cognitive stimulation and warm and sensitive parent-child interactions have been linked with positive child
outcomes, such as school readiness, math and literacy development, interest in learning, and communication and problem solving skills (Landry, Smith, & Swank, 2006; Martin, Ryan, & Brooks-Gunn, 2013; Price, 2010; Tamis-LeMonda et al., 2014), while use of harsh discipline practices such as hitting or spanking predict more negative outcomes like aggression, delinquent and antisocial behavior, poorer mental health, and impairments in cognitive ability (Gershoff & Grogan-Kaylor, 2016). Harsh discipline also has the potential to escalate into more serious child abuse (Gershoff & Grogan-Kaylor, 2016; Zolotor et al., 2008).

Given their links with positive child outcomes, many interventions seek to promote positive parent-child interactions and reduce the frequency of negative ones. Studies examining whether parents’ cognitively stimulating and socioemotional interactions with children can be positively modified have demonstrated that these behaviors are indeed amenable to change (Benasich, Brooks-Gunn, & Clewell, 1992; Easterbrooks et al., 2013; Green, Tarte, Sanders, & Waller, 2016; Love et al., 2005; Olds, Henderson, Tatelbaum, & Chamberlin, 1988; Reese, Sparks, & Leyve, 2010). This has been shown across interventions implemented within a variety of settings, including in families’ homes, early care and education centers, juvenile justice facilities, and community centers. As explained above, however, parenting programs tend to experience difficulty recruiting and retaining low-income parent participants. Programs that reach these parents through ECE programs thus may hold more promise for effectively engaging parents because many low-income parents already use ECE and thus must regularly engage with their children’s ECE setting and providers. Studies using random assignment to evaluate interventions implemented in ECE settings have demonstrated success in promoting parents’ literacy activities with children (Love et al., 2005; Mayer, Kalil, Oreopoulos, & Gallegos, 2015; York & Loeb, 2014), their parenting quality with regard to increased supportiveness and lower
detachment (Love et al., 2005), and their involvement in school (Van Voorhis, Maier, Epstein, Lloyd, & Leung, 2013; York & Loeb, 2014). They have also been shown to be successful in reducing spanking incidence (Love et al., 2005).

**Mechanisms by Which Head Start Could Impact Parenting**

It is well established that ECE—including Head Start—can impact children’s early academically-related outcomes. This understanding is reflected in empirical literature linking ECE to children’s outcomes (e.g. Bassok, Fitzpatrick, Greenberg, & Loeb, 2016; Gordon, Colaner, Usdansky, & Melgar, 2013; Gormley, Gayer, Phillips, & Dawson, 2005; McCartney, Dearing, Taylor, & Bub, 2007; NICHD ECCRN, 2002; Yoshikawa et al., 2013) and in our substantial financial investment in public early childhood programs, including Head Start, public pre-kindergarten programs, and child care subsidies. Of these investments, Head Start is a particularly promising avenue for supporting low-income children, given that it is reserved for children whose families are low-income and given that it has been historically well funded. Specifically, approximately $8.9 billion was spend on Head Start in 2017, which represented 16% of the Administration for Children and Families budget in that year—third only to Temporary Assistance for Needy Families, 31%, and Foster Care and Permanency, 18% (Administration for Children and Families, 2017; Office of Legislative Affairs and Budget, 2019).

Much of the impact of Head Start on children’s development is likely direct—the instruction, structure, and classroom experiences that children receive through Head Start likely directly influences their academic and behavioral skill development. But part of Head Start’s impact on children could also operate through its impact on parenting behaviors known to support child development. In other words, Head Start could influence the ways in which parents
interact with their children, which could in turn contribute to positive child outcomes. The link between Head Start and parenting behaviors has been far less studied than the link between Head Start and child outcomes. Even in the literature that focuses on two-generation approaches to promoting both child and parent outcomes simultaneously, the focus is typically on finding ways to help parents enhance their human capital while simultaneously helping children build their own early skills, as opposed to fostering parenting behaviors known to promote positive child outcomes (Chase-Lansdale & Brooks-Gunn, 2014). Studying the specific link between Head Start and parenting outcomes is important for understanding the full potential for Head Start to benefit low-income children.

There are a number of reasons to hypothesize that Head Start, and ECE in general, might impact the way that parents engage with children at home. First, Head Start might impact parents through direct instruction and modeling. That is, Head Start providers might be able to model or explicitly teach parents about effective or otherwise positive parenting behaviors – or about child development – in ways that could enhance their parenting (Bandura, 1977; Brooks-Gunn, Berlin, & Fuligni, 2000; Howard & Brooks-Gunn, 2009; Lee, Brooks-Gunn, Han, Waldfogel, & Zhai, 2014; Rafferty, Griffin, & Robokos, 2010). For example, parents might witness providers doing cognitively enriching activities with children, such as reading, talking about numbers, telling stories, or doing art or science projects that parents may then engage in at home. Similarly, providers may suggest ideas (solicited or not) for activities parents could do with children to further their learning, which the parent might then implement with the child at home (McCartney et al., 2007). With regard to discipline, a provider might be able to suggest strategies to effectively address misbehavior through non-physical means or might model such strategies that
parents might then repeat, thus reducing the frequency with which parents use harsh discipline at home.

Second, Head Start could provide parents time to enhance their human capital. More specifically, Head Start provides parents with the time to spend on other pursuits, including work and education, both of which improve mothers’ human and social capital. Links between maternal education and income and more optimal parenting practices are well established (Conger & Conger, 2002; Dodge et al., 1994; Guryan et al., 2008; Hart & Risley, 1995; Kalil et al., 2012; NICHD ECCRN, 2004; Phillips, 2011; Rowe, 2018; Suizzo & Stapleton, 2007), and possibly arise due to education’s influence on parents’ knowledge, skills, abilities, and effectiveness (Mirowsky & Ross, 2003; Rowe, 2018). Thus, increases in maternal education and employment may also improve mothers’ interactions with their children (Augustine, 2014; Harding, 2015; Harding, Morris, & Hill, 2017).

Finally, Head Start’s effect on parenting might also derive indirectly from its effects on children. It is well documented that while parents can influence child development, children also elicit behaviors from parents by virtue of their behavior and skills (Ansari & Crosnoe, 2015; Belsky, 1984; Paschall & Mastergeorge, 2016; Sameroff & MacKenzie, 2003; Thomas & Chess, 1977; Yan & Ansari, 2016). It is possible that children gain skills through ECE that subsequently elicit different behaviors from parents. For example, to the extent that ECE enhances children’s cognitive skills (Barnett, 1995; Camilli, Vargas, Ryan, & Barnett, 2010; Duncan & Magnuson, 2013; NICHD ECCRN, 2002, 2006; Weiland & Yoshikawa, 2013; Yoshikawa et al., 2013), those gains may cause children to elicit more cognitively stimulating behaviors from parents—children might ask parents to read to them, count with them, tell them stories, talk about topics they learned about in child care, or ask more complex questions in general, all of which could
provoke parents to respond accordingly. Furthermore, as ECE helps children to become more adept readers, counters, and conversationalists, parents might simply enjoy reading, counting, and talking to them more and therefore do so more often (Gelber & Isen, 2013). ECE also has the potential to promote children’s socioemotional skills – such as lowering internalizing and externalizing behavior problems and increasing sociability (Clarke-Stewart et al., 2002; Peisner-Feinberg et al., 2001; Votruba-Drzal & Chase-Lansdale, 2004) – which could, in turn, elicit more sensitive and less harsh caregiving from parents. It is important to note, however, that the literature on socioemotional outcomes is mixed, with some studies finding negative associations between time in center-based care and children’s socioemotional wellbeing (Belsky, 2002; Belsky et al., 2007; Morrissey, 2009; NICHD ECCRN & Duncan, 2003); thus, it is also possible that more time in ECE promotes less sensitive and more harsh caregiving if children’s behavior problems increase, making them more difficult to manage. Either way, the impact of ECE on the home environment could stem from its effects – first – on children. This mechanism does not imply that the effects of ECE on the home environment do not benefit children in turn. Rather, it is possible that the effects of ECE on children’s outcomes produce synergistic effects on the quality of the home environment that further enhance children’s development.

Directly testing these proposed mechanisms is beyond the scope of the present studies. Instead, explanations of these mechanisms are offered as suggestions for how to think about how Head Start might operate to change parenting behaviors. Although they are not directly tested, these mechanisms are supported by the bioecological model of development, which posits that development is a reciprocal interaction between the child and his or her environment and that this development occurs in multiple, often overlapping systems (Bronfenbrenner & Morris, 2006). Parents are thus influenced by their interactions with their children’s ECE providers, the
context in which they are parenting (which encompasses their education and employment status), and by their interactions with their children (who are themselves also influenced by the same providers).

**Empirical Evidence for the Impact of Head Start on Parenting**

Although the majority of research on the effects of Head Start has focused on child outcomes, there is a small but informative body of research on the effect of Head Start on parenting behavior. Some of these studies have used non-experimental designs to examine the association between Head Start attendance and parenting outcomes, relative to other types of ECE. Studies using non-experimental designs have found that children who attended Head Start were less likely to have low access to learning materials compared to children in other arrangement types (Zhai, Waldfogel, & Brooks-Gunn, 2013) and were less likely to be spanked compared to children in parent care (Magnuson & Waldfogel, 2005) and compared to children in other center-based arrangements (Zhai et al., 2013).

While informative, a major problem with non-experimental studies is that without a control group, they cannot tell us about the causal effect of Head Start on parenting behavior. The primary source of information on the causal effect of Head Start comes from the Head Start Impact Study (HSIS), which is the result of a congressional mandate directing the department of Health and Human Services (HHS) to determine whether Head Start was effective. In the HSIS, newly entering 3- and 4-year-old eligible children were randomly assigned for the 2002-2003 school year to either a Head Start group that had access to Head Start services, or to a control group that did not have access to Head Start but could enroll in other ECE programs or services if their parents chose to do so (Puma et al., 2010). This means that the “treatment effect” in the HSIS represents a comparison between Head Start and children’s local alternatives, including
parent care. Importantly, local Head Start programs are often overenrolled and thus choose children at random to receive Head Start services. Families who apply for but do not get a slot in Head Start are thus left to choose an alternative arrangement of their own—parents might choose to stay at home with their children rather than pay for care, leave the child with a relative or friend, or find an affordable alternative arrangement, either at a provider’s home or in a center.

Thus, the HSIS’s comparison between children randomly assigned to Head Start and children randomly assigned to not receive Head Start mimics real life distinctions between a set of low-income children and families who all desired Head Start services at some point. Results from the original HSIS evaluation and from subsequent studies using the same data thus provide meaningful information on parents in Head Start.

Studies using data from the HSIS have shown that parents in the 3-year-old cohort randomly assigned to the treatment condition were more likely (by seven percentage points) to engage in cognitively stimulating activities such as frequently reading to children and participating in culturally enriching activities with them than parents in the control condition, with effect sizes of 0.14-0.18 (Ansari et al., 2016; Puma et al., 2010). Head Start assignment in the 3-year-old cohort also had an effect on the amount of math-related activities parents did with children, such as playing math related games (Gelber & Isen, 2013). The HSIS final report did not find gains in parents’ reading, math activities, or culturally enriching activities for the 4-year-old cohort. In a study combining the 3- and 4-year-old cohorts, however, Head Start parents were found to read for longer (18.71 more minutes per week) and more frequently with children than control parents (Gelber & Isen, 2013).

Studies of Head Start examining parents’ disciplinary interactions with children have found that parents in the 3-year-old cohort randomly assigned to Head Start were seven
percentage points (effect size = -0.14) less likely to spank children compared to parents in the control condition (Puma et al., 2010). Parents in the 4-year-old cohort were no more or less likely to spank after the HS year, though they were less likely to give their children time outs than parents in the control condition after one year of Head Start. As noted in the HSIS final impact report, findings with regard to time out are worth noting, but the interpretation of changes in time out is not clear—decreases in time out could be due to positive changes in children’s behavior or to parents shifting to other, less positive, disciplinary strategies (though no increases in spanking were found; Puma et al., 2010).

**Variation in the Effect of Head Start Across Sites**

Despite the important past contributions of research demonstrating positive changes in parenting behavior as the result of Head Start, this body of work leaves a major question about the impact of Head Start on parenting unanswered. Specifically, this literature almost exclusively examines the average effect of either being randomly assigned to Head Start or of attending Head Start, ignoring potential heterogeneity of treatment effects across different Head Start centers. Exploring the extent to which centers vary in their effectiveness is important because we know that not all Head Start centers are the same. Although Head Start has a central funding source and mission, individual Head Start centers vary in their management, how they choose to operate, the parenting programming they offer, their classroom quality, the curricula they use, and in the types of families they serve. Thus, a focus on the average effect of Head Start could mask variation in effects across programs that could help to elucidate why some programs are more effective than others (Bloom, Raudenbush, Weiss, & Porter, 2017; Bloom & Weiland, 2015; Weiss et al., 2017; Weiss, Bloom, & Brock, 2014).
Indeed, there is growing acknowledgement among researchers interested in interventions that we need a better understanding of “for whom” and “under what conditions” interventions are most effective (Reardon & Stuart, 2017). Although this has been argued conceptually for decades, evaluation research and public policy analysis to date has largely focused on the average effects of programs while paying very little attention to exploring and explaining variation (Weiss et al., 2014). In recognition of this scarcity, there has been a recent increase in efforts to quantify and to explore reasons for variation in the effects of interventions where the same program is being implemented across multiple sites—as in multi-site RCTs (Bloom et al., 2017; Sabol et al., 2019; Walters, 2015; Weiss et al., 2017). Recent investigations into variation in Head Start’s impact on child outcomes using the HSIS have demonstrated that there is meaningful variation in Head Start’s effects on children’s cognitive and socioemotional outcomes (Bloom et al., 2015; Walters, 2015). For instance, Bloom and Weiland (2015) found that Head Start centers varied significantly in the degree to which they improved children’s receptive vocabulary, early reading, oral comprehension, externalizing behavior, and self-regulation skills. Using composite measures of cognitive and noncognitive skills, Walters (2015) found significant treatment effect variation across Head Start centers for both domains.

Once it has been determined that meaningful variation across programs exists—meaning that there is more variation in outcomes across programs than would be expected by chance alone—it is possible to identify sources of that variation. The premise here is that if there is variation in a given program feature across programs, and that feature is an important predictor of Head Start’s effectiveness, then there will also be variation in treatment effects across programs. Evaluations seeking to uncover sources of variation in Head Start treatment effects on child outcomes point to a number of such features. Factors that have been identified as
explaining why some programs are more effective than others at improving children’s academic and behavioral outcomes include variation across centers in the representation of dual-language learners with low initial test scores (Bloom & Weiland, 2015), in centers’ urbanicity (McCoy, Morris, Connors, Gomez, & Yoshikawa, 2016), in center characteristics such as whether centers offered full-day service, opportunities to observe or practice parenting skills, or home visits (Feller, Grindal, Miratrix, & Page, 2016; Walters, 2015), in the type and quality of care that children in the control condition received (Feller et al., 2016; Friedman-Krauss, Connors, & Morris, 2017), and in state child care licensing regulations (Connors & Friedman-Krauss, 2017).

To date, all evaluations seeking to quantify and explain variation in Head Start effectiveness has been limited to explorations of variation in children’s outcomes. With regard to Head Start’s potential to impact parenting behaviors, no prior study has gone beyond examining subgroup differences to more thoroughly quantify and explore reasons for variation in Head Start treatment effects on parenting outcomes. However, we might expect there to be variation across programs in the extent to which they are effective at improving parents’ cognitively stimulating and socioemotional interactions with children for many of the same reasons we would expect effects to vary for child outcomes—management, quality, programming, operational decisions, and the types of families served all vary across Head Start programs in ways that could produce variation in program effectiveness for parents. Indeed, in their conceptual framework for studying the sources of variation in program effects, Weiss and colleagues (2014) posit three factors that might moderate the causal relationship between a particular intervention’s treatment and the program’s effects: characteristics of the organization, “client characteristics,” which are characteristics of the individuals participating in the intervention, and the treatment contrast—the difference between what the treatment and control group of an RCT experience. This dissertation
tests whether factors from the first two of these sources—characteristics of Head Start centers and characteristics of the families they serve—help to explain why some centers are more effective than others at improving parenting outcomes.

**Potential Sources of Head Start Variation in Parenting Outcomes: Center Characteristics**

An important potential source of variation in Head Start effects across centers is characteristics of centers themselves. Although Head Start operates through a central funding source and is highly regulated, individual centers vary across a host of domains—they vary with regard to how they choose to implement Head Start standards or initiatives, in their quality, their programming, their delivery, and in other relevant ways. This variation in center characteristics translates to real variation in the treatment that is received by different families. Thus, Head Start centers with certain characteristics might make them more likely to impact parenting practices than centers without those characteristics. To the extent that this is true, variation in a given center characteristic that is highly predictive of Head Start’s effectiveness should produce variation in effects across Head Start centers.

Returning now to the mechanisms by which Head Start might impact parenting outlined above, Head Start could impact parents through direct instruction and modeling, through improvements in parents’ human capital, and through its effects on children. The focus of this dissertation is on exploring moderators of Head Start’s effects on parenting outcomes that are related to the direct instruction and modeling that Head Start centers do. I chose to focus on this first mechanism for two reasons. First, this mechanism is most closely aligned with Head Start’s historical goal of empowering parents to be active participants in their children’s learning and their recent efforts to promote family engagement in Head Start through its Parent, Family, and Community Engagement Framework (Administration for Children and Families, 2018). Second,
of the proposed mechanisms by which Head Start could impact parenting, direct instruction and modeling is arguably the most direct route to changing parenting—both human capital enhancement and effects on children would operate to influence parenting indirectly through other channels. Thus, factors related to direct instruction and modeling may have the greatest potential to induce positive changes in parenting.

Two center characteristics that might influence a center’s ability to provide direct instruction and effective modeling, and thus influence a center’s effectiveness at improving parenting outcomes, are parent outreach and program quality. Parent outreach includes activities that occur inside and outside of Head Start centers that are aimed at engaging parents in their children’s development. Activities done inside the center include workshops, formal meetings between teachers and parents, and even making informal suggestions to parents during pick-up or drop-off. Activities outside the center include home-visiting services. During these times, providers have the opportunity to give parents information about their own child’s progress or about child development more broadly, suggestions about activities they can do with their children at home, and suggestions about strategies to handle kids’ misbehavior without becoming physical. Providers can also model effective and appropriate cognitive stimulation and disciplinary techniques for parents during these times. Indeed, parent outreach done in ECE centers has been linked with improvements in parenting behaviors (Grindal et al., 2016; Sheldon & Epstein, 2005), and home visiting services provided through Head Start are effective at promoting positive parenting behaviors (Love et al., 2005). Importantly, centers likely vary in the degree to which they perform in-center outreach activities. And, although Head Start centers are required to provide two home visits to families per year, they can provide more if they choose, producing variability in the number of home visits that are offered. To the extent that
these practices are linked with improvements in parents’ outcomes, variability in outreach activities performed by centers likely produces variability in centers’ effectiveness.

A second important center characteristic that may help determine the extent to which centers impact parenting is classroom quality. Dimensions of ECE quality fall into two categories: process quality—features pertaining to the quality of teacher-child interactions—and structural quality—features related to the setting structure or requirements such as group size, ratio, or teacher qualifications. High-quality ECE settings, especially those characterized by warm and responsive interactions between children and providers and that provide access to a range of early learning materials such as books, blocks, and sensory experiences (Mashburn et al., 2008; Yoshikawa et al., 2013), are likely better able to promote positive parenting outcomes through direct instruction and modeling than lower quality centers. Providers in these settings are likely more effective at and better equipped to communicate and engage with parents and to model enriching and sensitive interactions with children than providers at lower quality centers. For example, a parent can only learn more stimulating and sensitive behaviors from providers if the providers’ behavior with children is itself stimulating and sensitive. It is also more likely that providers in high-quality settings are more willing and able to provide childrearing guidance and advice and to impart knowledge of child development than those from lower quality centers.

Indeed, there is some evidence that higher quality child care is associated with improved maternal responsiveness, cognitive stimulation, and increased availability of learning materials in the home, particularly for poorer families (McCartney et al., 2007). Furthermore, classroom quality has been shown to vary across Head Start centers (Aikens, Bush, Gleason, Malone, & Tarullo, 2016; Alamillo et al., 2018; Puma et al., 2010). Thus, to the extent that classroom quality is indicative of a center’s ability to provide direct instruction and modeling to parents,
quality differences across Head Start centers could result in differences in effectiveness across centers.

**Potential Sources of Head Start Variation in Parenting Outcomes: Family Characteristics**

Head Start centers do not just differ in their program characteristics. They also differ in that they each serve a different population of families, who themselves likely differ in their responsiveness to the direct instruction and modeling that Head Start provides. If Head Start benefits families with a given characteristic more than families without that characteristic, then centers serving families with high percentages of families with that characteristic should be more effective at improving parenting outcomes than centers serving lower percentages of families with that characteristic. Four characteristics of families that could potentially make them more or less able or likely to benefit from Head Start’s direct instruction and modeling include families with multiple economic risk factors, families with an immigrant parent, families in which parents attended Head Start as children, and families with low levels of parental cognitive stimulation when they began Head Start.

All families served by Head Start are low-income and therefore at risk for poorer parenting outcomes (Bradley, Corwyn, Pipes McAdoo, & García Coll, 2001; Dodge et al., 1994; Gershoff & Grogan-Kaylor, 2016; Guryan et al., 2008; Kalil et al., 2012; Sayer, Gauthier, & Furstenberg, 2004). However, families experiencing multiple economic risk factors typically experience more financial and emotional strain than other low-income families (Butterworth, Rodgers, & Windsor, 2009; Mcloyd, 1990), factors strongly associated with lower levels of parental cognitive stimulation and higher levels of harsh punishment (Guryan et al., 2008; Kalil et al., 2012; Ryan, Kalil, Ziol-Guest, & Padilla, 2016; Sayer, Gauthier, et al., 2004; Straus & Paschall, 2009; Straus & Stewart, 1999). Thus, these parents might be particularly likely to
benefit from the direct instruction and modeling that Head Start centers provide because they arguably need the most support when it comes to learning about effective parenting strategies.

Similarly, immigrant parents face multiple risk factors to those parenting behaviors known to promote positive child outcomes. Specifically, immigrant parents tend to have fewer financial and educational resources than non-immigrant parents (Lopez & Velasco, 2011; Park & McHugh, 2014; Pew Research Center, 2015), which, as noted above, could make them more likely than other parents to benefit from the positive suggestions and modeling that Head Start providers provide. As a group, they also tend to have high expectations for their children and to place a high value on formal education (Goldenberg, Gallimore, Reese, & Garnier, 2001; Hao & Bonstead-Bruns, 1998), which could perhaps further signal receptiveness to Head Start providers’ suggestions and modeled behavior. On the other hand, immigrant parents also tend to have more limited exposure to English than non-immigrant parents (Gambino, Acosta, & Grieco, 2014; Park & McHugh, 2014), which, depending on the language match between parents and providers, could hinder their ability to communicate with providers and thus impede their ability to benefit from the direct instruction and modeling provided by their Head Start centers.

Families in which both the mother and child participated in Head Start as children—known as multigenerational Head Start families—may also experience different benefits with regard to their parenting practices than single-generation families. There are reasons to posit that multigenerational Head Start parents might benefit both more and less from the program than single-generation Head Start parents. Mothers who went to Head Start themselves were low-income as children and have not escaped poverty as adults, signaling a substantial risk to positive parenting practices. However, many mothers from single-generation Head Start families were likely also disadvantaged as children (Aaronson & Mazumder, 2008; Black & Devereaux, 2011;
Isaacs, Sawhill, & Haskins, 2008; Lindahl, Palme, Massih, & Sjögren, 2015), so experiencing poverty as a child does not necessarily distinguish multi- from single-generation families. Moreover, parents who have participated in Head Start as children themselves may be more familiar with and receptive to Head Start’s messaging and programming, and may place greater trust in Head Start as an institution than other families. This greater familiarity, receptivity, and trust may make multigenerational parents more responsive to Head Start providers’ direct instruction and modeling. Thus, multigenerational Head Start parents’ parenting practices may actually benefit more from sending their children to Head Start than parents with more limited prior exposure to and trust in the program.

Finally, there is wide variation in parenting attitudes and behaviors, and Head Start parents enter the program with their own parenting practices and routines already established. Parents who enter the program engaging in lower levels of cognitive stimulation theoretically have more “room to grow” from the program and thus may benefit more from Head Start providers’ suggestions and modeling than parents who enter Head Start already practicing high levels of positive parenting behaviors. For instance, parents with low initial cognitive stimulation levels may be unfamiliar with the benefits of daily reading or of practicing math concepts with young children such that once they learn about them from a Head Start provider, they can make a bigger improvement in those parenting behaviors than families who were already practicing them. On the other hand, parents who begin the Head Start year with the highest levels of positive parenting practices have already demonstrated high levels of motivation and commitment to their children’s development. These parents may thus be more receptive to Head Start’s direct instruction and modeling, leading them to respond more strongly to Head Start’s
messaging with further improvements to their baseline parenting practices than parents with lower levels of baseline cognitive stimulation.

**Using the Head Start Impact Study to Study Treatment Effect Variation**

When the same program or intervention is implemented and then evaluated across many different sites, as in multi-site RCTs, it is possible to examine variation in treatment effects across the program sites. The HSIS, a multi-site RCT of Head Start, is the ideal dataset with which to answer questions about variation in Head Start effects across centers. Importantly, randomization in the HSIS occurred at the center-level, 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 thought of as a composite of approximately 300 randomized controlled trials (RCTs), each conducted at a single site. Use of the HSIS dataset allows me to effectively combine estimates from each site to estimate both an average causal effect and also a standard deviation for that effect that represents how much those effects vary across Head Start centers.

**Threats to Causal Inference in the Head Start Impact Study**

Although the randomized controlled nature of the HSIS is vital for being able to produce causal estimates of the effect of Head Start on parenting behaviors, there is one feature of the study that must be kept in mind when discussing causal effects. Specifically, there was significant non-compliance to random assignment in the study, meaning that some children randomly assigned to the treatment condition did not attend Head Start and likewise, some children randomly assigned to the control condition nevertheless managed to enroll in Head Start. About 14-20% of sample members did not comply with their random assignment, depending on children’s cohort and random assignment result.
The substantial non-compliance present in the HSIS means that it is necessary to estimate the impact of random assignment separately from the impact of participation in Head Start. Both estimates are useful. The effect of random assignment impact 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). By estimating both effects, this dissertation aims to inform both policy and program implementation.

The Present Studies

This dissertation extends prior literature on the effect of Head Start on changes in parenting behaviors by considering variation in those effects across Head Start centers in a series of three studies. Study 1 replicates past work by examining the average effect of both Head Start random assignment and participation on parents’ cognitively stimulating and disciplinary interactions with their children. Study 1 also extends this past work by quantifying the cross-site variation in treatment effects on parenting outcomes and explicitly testing whether that variation is statistically different from zero, which no other study has done. After demonstrating that Head Start has a significant effect on multiple parenting behaviors and that many of those effects vary significantly across Head Start centers, Studies 2 and 3 examine whether various theoretically important center and family characteristics explain why centers vary in their treatment effects. Specifically, Study 2 asks whether programs that do more parent outreach activities or that have higher classroom quality are more effective at changing parenting outcomes than programs that do fewer parent outreach activities or are lower quality. Study 3 asks whether four meaningful family characteristics—family economic risk, maternal immigrant status, multigenerational Head
Start status, and parents’ pre-academic stimulation—help predict which centers are most effective at changing parenting behaviors. Together, these studies will help to provide a better understanding of the center and family characteristics of programs that are most successful at promoting positive parenting practices among the families they serve. This understanding could help inform Head Start decision makers about how to best target efforts aimed at achieving Head Start’s goal of empowering parents to actively engage in their children’s development.
CHAPTER II: VARIATION IN HEAD START PROGRAM EFFECTS ON PARENTS’ INTERACTIONS WITH CHILDREN

Since its inception in 1965, Head Start has served more than 35 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 (Administration for Children and Families, 2017). The main goal of Head Start is to improve the school readiness of children from low-income families. This goal is primarily achieved by providing free early care and education (ECE) for the families it serves, however, Head Start also adheres to a “whole child” model. Thus, in addition to providing ECE services, Head Start also provides medical, dental, nutritional, and mental health care, and actively encourages parents’ involvement in their children’s learning (Puma et al., 2010; Zigler & Styfco, 2004). The aim of involving parents in their children’s development is reflected in Head Start’s historical focus on promoting parents’ wellbeing and interactions with children (Puma et al., 2010; Zigler & Valentine, 1979) as well as the newly updated Head Start performance standards, which emphasize efforts to involve parents and families in their children’s early learning through its Parent, Family, and Community Engagement Framework (Administration for Children and Families, 2018).

Although the majority of research on Head Start’s effectiveness has been focused on determining whether Head Start has any immediate or lasting effect on low-income children’s school readiness, there has also been some research into whether Head Start has an effect on parents’ interactions with their children. The parenting behaviors most of interest to Head Start are those that are known to promote children’s school readiness, including cognitively stimulating and warm, sensitive interactions (Beck, 1999; Brooks-Gunn & Markman, 2005; Dodge et al., 1994; Gershoff, 2002; Goodman et al., 2011; Hubbs-Tait et al., 2002; Landry, Smith, & Swank, 2003; Putnam, Sanson, & Rothbart, 2002; Raikes et al., 2006; Yeung,
Sandberg, Davis-Kean, & Hofferth, 2001). Research examining the potential for Head Start to promote positive parenting behaviors has found that participation in Head Start is associated with small to moderate increases in parent reading and math-related activities and lower levels of harsh punishment such as spanking after one year of the program (Ansari et al., 2016; Gelber & Isen, 2013; Magnuson & Waldfogel, 2005; Puma et al., 2010; Zhai et al., 2013). While informative, prior research examining Head Start’s effects on parenting behaviors has largely focused on the average effect of Head Start on parents. This focus is problematic because solely examining Head Start’s average effect masks potential heterogeneity of effects across programs. Knowing whether and how much Head Start effects vary across programs is an important first step in elucidating optimal program practices and conditions.

The present study first estimates the average effect of Head Start on parents’ cognitively stimulating and disciplinary interactions with children, an inquiry that prior studies have made to varying degrees. Next, this study addresses the identified gaps in the literature by quantifying how much Head Start program effects on parents’ behavior vary across programs and explicitly testing whether that variation is meaningfully different from zero. Quantifying treatment effect variation in this way involves describing the extent to which program impacts are similar or different across Head Start sites (Sabol et al., 2019). This study lays the groundwork for exploring the potential reasons for variation in Head Start treatment effects across programs and thus is an important first step in understanding Head Start’s full impact on parenting behavior.

**Variation in Head Start Program Effects on Parenting Behaviors**

Although Head Start centers operate under a central funding source, 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 many ways, for example in their program curricula, the parenting programming they offer, their teachers’ qualifications, or even in the overall climate and attitude towards parents. It is thus important to understand how much centers vary in their effectiveness so that we can begin to understand what determines program effectiveness, and ultimately, improve services and outcomes for children and families.

In recognition of the need to study treatment effect variation, new and innovative methodologies have been developed that make it possible to more accurately quantify treatment effect variation using multi-site randomized controlled trials (RCTs) (Bloom et al., 2017; Raudenbush & Bloom, 2015; Weiss et al., 2017). There have been some recent efforts to quantify variation in Head Start’s treatment effects on children’s outcomes using these methodologies (Bloom & Weiland, 2015; Walters, 2015). These studies found that the originally reported small impacts on children’s academic and behavioral skills from the Head Start Impact Study (HSIS) were masking large variation in effects. This variation indicates that Head Start programs vary significantly in their effectiveness, with some programs more and some programs less effective than their local alternatives at promoting children’s school readiness (Bloom & Weiland, 2015; Walters, 2015).

Results from these studies provide us with important information about the potential for Head Start centers to vary in their effectiveness. However, all prior evaluations of heterogeneous treatment impacts across Head Start centers have solely examined children’s outcomes. By contrast, we have a very limited understanding of for whom and under what conditions Head Start centers are best able to promote parenting behaviors in the home that are known to impact children’s school readiness outcomes. Such parenting behaviors include increasing time spent
reading books and engaging in other cognitively stimulating activities, as well as reducing the incidence and frequency of harsh disciplinary practices.

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 with their children and on parents’ disciplinary interactions with their children (Ansari et al., 2016; Gelber & Isen, 2013; Puma et al., 2010). However, significant average effects do not necessarily signal that treatment effects vary across programs. Similarly, null average effects do not mean that there is no variation in impacts. Indeed, in a study that estimated both average effects and treatment effect variation using data from 16 large multisite RCTs of various education and workforce development interventions, Weiss and colleagues (2017) found that treatment impacts for these interventions fell into four categories. These categories illuminate the possible patterns for multisite RCT treatment impacts more broadly. 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 across sites. 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 programs. Third, treatment effects could be consistently positive, meaning that the average effect is positive and there is little variation in that positive effect. Fourth, there could be a significant positive impact on average with substantial cross-site variation, where most or all sites produce positive effects with widely varying magnitudes.

The present study considers two types of parenting outcomes: parents’ cognitively stimulating and disciplinary interactions with their children. Considering the four potential patterns of results outlined above for multisite RCTs, while also considering prior Head Start
studies demonstrating average positive effects for both of these domains, there are a number of potential patterns that could emerge in the present investigation. First, it is possible that Head Start has a similarly sized positive effect on both cognitive stimulation and disciplinary parenting outcomes, but that neither effect varies significantly across sites. In this scenario, Head Start is consistently effective. This pattern would signal that Head Start is a very centralized program that is likely implemented similarly across sites and affects different types of families similarly. A second possibility is that Head Start could have a similarly sized positive effect on both types of outcomes, and that effect varies across programs for one or both types of behaviors. This pattern would suggest that Head Start is effective on average for both types of outcomes, but that there is variation such that sites vary in their promotion of one or both outcomes. If only one outcome type varies, this would suggest that Head Start is consistently better at promoting one of the outcomes over the other, for which effects are more variable. If both outcome types vary, it would suggest that Head Start is more varied than consistent in its ability to promote positive parenting outcomes. Third, it is possible that Head Start effects are stronger for one type of outcome than the other, but that variation is stronger for the outcome with weaker average effects. This pattern would signal that Head Start is consistently good at promoting one outcome, and sometimes – but less consistently – good at promoting the other. Finally, it is possible that Head Start effects are on average stronger for one type of outcome than the other, and that variation is also wider for the outcome with the stronger main effects. This pattern would indicate that Head Start is not very effective at promoting the outcome with the weaker effects and that its greater effectiveness at promoting the other outcome still varies at the site level. Studying both the average effects and variation in effects across sites therefore provides
considerably more information about how Head Start is operating than focusing solely on the average effect.

**Studying Treatment Effect Variation in the HSIS**

Two features of the HSIS make it the ideal dataset with which to answer questions about variation in Head Start treatment effects. First, children and families were randomly assigned to a treatment group, who had access to Head Start services, or to a control group, who did not have access to Head Start but could enroll in other programs. This randomized controlled design is essential for being able to obtain causal estimates of the effect of Head Start. Second, randomization occurred at the center-level, 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 second feature means that the HSIS can be considered a composite of over 300 RCTs, each conducted at a single site. This design allows me to combine estimates across sites to estimate both an average causal effect and a standard deviation for that effect that quantifies how much those effects vary across Head Start centers.

Although the ideal dataset for answering questions about Head Start treatment impact heterogeneity, there was significant non-compliance to random assignment in the study. Specifically, some children randomly assigned to the treatment condition did not attend Head Start (“no shows” account for about 15% and 20% for the 3-year-old and 4-year-old cohorts, respectively) and likewise, some children randomly assigned to the control condition nevertheless managed to enroll in Head Start (“crossovers” account for about 17% and 14% for the 3-year-old and 4-year-old cohorts, respectively). It is thus 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 non-compliance.

The Present Study

Using data from the HSIS, the present study first replicates prior reports of the average effect of both Head Start random assignment and enrollment on parents’ cognitively stimulating and disciplinary interactions with their children. Next, in an extension of prior work, the present study takes an important first step in understanding Head Start program effect variation by asking whether and how much Head Start centers vary in their effectiveness at improving parents’ cognitively stimulating and disciplinary interactions with their children after one year of Head Start. I do this by quantifying the heterogeneity in Head Start treatment effects on parenting outcomes and testing whether that variation is significantly different from zero. Here, testing whether variation is statistically different from zero means testing the null hypothesis that the variation in effects across programs is zero.

Methods

Data and Sample

The HSIS is a nationally representative sample of 383 randomly selected Head Start centers and 4,667 newly entering, eligible three- and four-year-old children. The study began in 2002 and continued through 2006, when children ended first grade. Following others (Bloom & Weiland, 2015; McCoy et al., 2016), the current study combined the three- and four-year cohorts and examined short-term parenting outcomes after one 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 to a control group that did not have access to Head Start (but could enroll in other programs). Importantly, randomization occurred at the center-level, 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 (with some exceptions for very small Head Start centers, for which centers were combined into center-groups; in these cases randomization occurred at the center-group level; Puma et al., 2010). Thus, the HSIS can be considered a composite of 383 randomized experiments, each conducted at a single site (Bloom et al., 2017; Bloom & Weiland, 2015).

Separate analytic samples were created for each parenting outcome. Each sample was restricted to sample members who had valid data for the parenting outcome under investigation. Following others (Bloom et al., 2017; McCoy et al., 2016), missing predictor variable data (e.g. missing data on covariates and baseline parenting variables) were addressed through single replication of a multiple imputation model. Imputation was done 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, centers were then dropped if after restricting to those with non-missing outcome data they either had 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). Centers were also dropped 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 296-297 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 harsh discipline. Three cognitive stimulation outcomes were included in this study: whether parents read to children daily as well as the frequency of their literacy- and math-related activities with children. Two harsh discipline outcomes were included: whether parents spanked their children in the past week and the number of times they did so.

**Spring cognitive stimulation.** Cognitive stimulation in the Spring consisted of three items: whether parents read to children daily as well as the frequency of their literacy- and math-related activities with children.

Parents reported on the item “How many times have you or someone in your family read to [CHILD] in the past week?” Possible responses for the reading frequency variable range from 1 (not at all) to 4 (every day). This variable was used to create a dichotomous reading variable indicating whether parents reported reading to children daily or not.

Parents also answered questions pertaining to how often they engaged in other enriching activities with children in a typical week. In the Spring, these activities fell into two categories: literacy- and math-related activities. Parents responded to the item “During the day, how often do you or someone in your household do each of the following reading and language activities with [CHILD]?” for 12 activities such as work on learning the names of the letters, discuss new words, and have child practice writing or spelling his/her name (see Appendix for full list of items). Responses ranged from 1 (Never) to 6 (Every day). The literacy activities composite is an average of parents’ responses for 13 items: these 12 literacy-related items, plus the non-dichotomized reading frequency variable described above (which was re-scaled to be on the same 1-6 scale as the other enrichment items) (Cronbach’s alpha = .89).

Parents also reported on the item “During the day, how often do you or someone in your household do each of the following math activities with [CHILD]?” for 8 activities such as count
out loud, work or shape blocks, and play math-related games (see Appendix for full list of items). Responses ranged from 1 (Never) to 6 (Every day). The math activities variable is an average of parents’ responses for these 8 items (Cronbach’s alpha = .81).

**Fall cognitive stimulation.** Cognitive stimulation in the Fall consisted of two items: whether parents read to children daily as well as the frequency of their overall cognitively stimulating activities with children. These variables served as covariates in models predicting parents’ Spring cognitive stimulation outcomes (see analytic strategy below).

As in the Spring, parents reported on the item “How many times have you or someone in your family read to [CHILD] in the past week?” Possible responses for the reading frequency variable range from 1 (not at all) to 4 (every day). This variable was used to create a dichotomous reading variable indicating whether parents reported reading to children daily or not.

It was not possible to create two separate literacy and math activities variables in the Fall because the HSIS did not ask about enough distinct math-related activities in the Fall. Thus, a broader “cognitive stimulation” composite variable was created for the Fall to serve as a covariate in models predicting both literacy and math activities in the Spring (see analytic strategy below). In the Fall, parents reported on the item “In the past week, have you or someone in your family done the following things with [CHILD]?” for 14 activities such as told him/her a story, helped him/her learn the names of letters, words, or numbers, practiced writing the letters of the alphabet with him/her (see Appendix for full list of items). Possible responses included zero (did not do activity), one (did activity one or two times), or two (did activity three or more times). The Fall cognitive stimulation variable is an average of parents’ responses for 15 items: the 14 enrichment activities, plus the non-dichotomized reading frequency variable described
above (which was re-scaled to be on the same 0-2 scale as the other fall enrichment items) (Cronbach’s alpha = .78).

**Harsh discipline.** In the Fall and the Spring, parents reported on the item “Sometimes children mind pretty well and sometimes they don’t. Have you spanked [CHILD] in the past week for not minding?” Spanking occurrence is a dummy variable indicating whether parents reported spanking their children or not. If parents indicated that they had spanked their child in the past week, they reported on the item “About how many times in the past week?” Spanking frequency is a variable indicating how many times parents reported spanking children in the past week.

**Treatment status.** Treatment status for models estimating the impact of random assignment to Head Start represents whether the child was randomly assigned to Head Start (1) or not (0). Treatment status for models estimating the impact of Head Start participation reflect whether the child was enrolled in Head Start (1) or not (0) during the Fall of 2002.

**Covariates.** Despite random assignment procedures, it is still important to control for baseline characteristics of families that may have differed between treatment and control groups due to differential response rates between groups (Puma et al., 2010). In each model, covariates included those used in the original HSIS final report analyses and were all taken from parents’ reports at baseline. These included child gender, child race/ethnicity, home language, mother's age, mother's education, mother's marital status, whether mother was a teenager at the time of child's birth, mother's immigrant status, and whether child lived with both biological parents. All models also controlled for the child’s age cohort. As discussed in the analytic strategy below, the appropriate parenting measure from the Fall also served as a covariate in each model predicting the corresponding parenting outcome in the Spring.
Analytic Strategy

To quantify variation in Head Start’s effects on parenting outcomes, I employ recently developed approaches designed for use with multi-site randomized controlled trials (Bloom et al., 2017; Bloom & Weiland, 2015; Raudenbush & Bloom, 2015; Raudenbush, Reardon, & Nomi, 2012) to 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 reduces the influence of unmeasured variables statistically by controlling 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; NICHD ECCRN & Duncan, 2003; Votruba-Drzal & Chase-Lansdale, 2004). Given random assignment, these unmeasured differences are less of a concern than in non-experimental studies, but nonetheless including these lagged dependent variables increase confidence in estimates (Puma et al., 2010). Regression coefficients are thus interpreted as the effect of random assignment or enrollment on changes in the parenting measure being predicted (Kessler & Greenberg, 1981).

Following others, (e.g. Bloom & Weiland, 2015a; Puma et al., 2010), for all studies I consider p-values less than .10 as statistically significant given that this matches the threshold used in the original HSIS final impact report.

Effect of random assignment. To estimate the effect of random assignment to Head Start – also referred to as 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). There are two features of this approach worth elaborating upon, given their departure from more traditional approaches. First, this two-level random coefficients model specifies a fixed intercept and a random coefficient for each site (called a FIRC approach; Sabol et al., 2019). The fixed intercept represents the average outcome for the control group; specifying a fixed intercept eliminates biases resulting from systematic and non-systematic differences across Head Start centers (Bloom & Weiland, 2015; Bloom et al., 2017; McCoy et al., 2016). The random coefficient represents the average treatment effect—specifying a random coefficient allows treatment effects to vary across Head Start sites. Second, the model specifies separate individual-level residual outcome variances for the treatment group and the control group. This 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).

Specifically, I estimate the following model:

**Level 1: Parents**

\[ Parenting_{ijT2} = \alpha_j + \beta_j RA + \lambda_i Parenting_{ijT1} + \sum_{k=1}^{K} \pi_k X_{kij} + \epsilon_{ij} \]  
**[Eq. 1]**

**Level 2: Head Start Centers**

\[ \alpha_j = \alpha_j \]  
**[Eq. 2]**

\[ \beta_j = \beta_0 + r_j \]  
**[Eq. 3]**

where \( i \) indexes individual parents, \( j \) indexes individual Head Start centers, \( T1 \) refers to Time 1 or the beginning of the Head Start year, and \( T2 \) refers to Time 2 or the end of the Head Start year. \( Parenting_{ij} \) represents each parenting outcome; \( RA_{ij} \) is an indicator representing random assignment status; \( X_{kij} \) represents each baseline covariate \( k \); and \( \epsilon_{ij} \) is a random error that varies.
independently across individuals with a mean of zero and variances of $\sigma_T^2$ and $\sigma_C^2$ for treatment and control group members, respectively. At level 2, $\alpha_j$ is a fixed intercept representing the mean control group outcome for Head Start center $j$; and $\beta_j$ is a slope that varies randomly across centers and represents the average effect of random assignment for Head Start center $j$; $\beta_o$ is the cross-site grand mean effect of random assignment to Head Start (this is the mean of all the center mean effects); and $r_j$ is a random error that varies across sites with a mean of zero and a variance of $\tau_{ITT}^2$.

$\beta_o$ and $\tau_{ITT}^2$ are the main parameters of interest for the present investigation given that $\beta_o$ estimates the average effect of Head Start random assignment, and $\tau_{ITT}^2$ indicates how much that effect varies. The statistical significance of $\beta_o$ was tested using the corresponding $z$ statistic from the 2-level mixed model, which was obtained using the MIXED command in Stata 15.0. The statistical significance of $\tau_{ITT}^2$ was tested using the conventional Q-statistic for random-effects meta-analysis (Bloom et al., 2017; Bloom & Weiland, 2015; Hedges & Olkin, 1985; Weiss et al., 2017). The Q-statistic has a $\chi^2$ distribution with J-1 degrees of freedom and tests the null hypothesis that $\tau_{ITT}^2 = 0$.

It is noteworthy that the approach taken here to estimate ITT effects is distinct from traditional methodological approaches that use ordinary least squares (OLS) regression with treatment-by-center interactions and center fixed effects. This more traditional OLS approach is problematic because the interacted fixed effect coefficients represent both cross-site variation in true program effects and random estimation error and thus often over-estimates treatment effect variation (Bloom & Weiland, 2015; Sabol et al., 2019). Using this FIRC approach instead makes it possible to estimate variation in true program effects, as opposed to program effect estimates (Bloom & Weiland, 2015).
Effect of enrollment. In the case of the HSIS, where there are both “no shows” and "cross overs,” the effect of Head Start enrollment represents a “Local Average Treatment Effect” (LATE), meaning that the average treatment effect is specific to compliers (Angrist, Imbens, & Rubin, 1996; Raudenbush et al., 2012).

In randomized controlled trials with considerable noncompliance, the method of instrumental variables (IVs), where treatment assignment is used as an “instrument” to identify the impact of participating in the program, is typically recommended (Angrist et al., 1996; Heckman & Robb, 1985; Raudenbush et al., 2012). However, typical IV models assume no site-level variation in treatment effects and so do not allow for estimation of variation in the causal effects of participation. Therefore, in the present investigation I follow an approach proposed by Raudenbush, Reardon, and Nomi (2012; “Option C”) and then expanded upon by Bloom and Weiland (2015) that allows for estimation of cross-site variation in enrollment effects, accounts for the nested structure of the data, and accounts for random assignment noncompliance. This approach involves two steps.

Step 1. In the first step, I apply two-stage least squares (2SLS) to a multiple instrumental variables model with $j$ Site x Treatment interactions serving as instruments in order to estimate the mean causal effect of Head Start enrollment for children from each Head Start center. The first stage of this model estimates the probability of enrolling in Head Start for each family based on whether they were randomly assigned to the treatment condition and all covariates. The second stage uses this predicted probability of enrolling to estimate the effect of Head Start enrollment. More specifically, I estimate the following two-stage model:
First-Stage Parent-Level:

\[ \text{Enrollment}_{ij} = \sum_{m=1}^{M} \theta_m \text{Center}_{mj} + \sum_{m=1}^{M} \gamma_m \text{RA}_{ij} \times \text{Center}_{mj} + \varphi_i \text{Parenting}_{T1} + \sum_{k=1}^{K} \psi_k X_{kij} + e_{1ij} \]  
[Eq. 4]

Second-Stage Parent-Level:

\[ \text{Parenting}_{T2ij} = \sum_{m=1}^{M} \phi_m \text{Center}_{mj} + \sum_{m=1}^{M} \delta_m \hat{\text{Enrollment}}_{ij} \times \text{Center}_{mj} + \partial_i \text{Parenting}_{T1} + \sum_{k=1}^{K} \eta_k X_{kij} + e_{2ij} \]  
[Eq. 5]

where \( \text{Enrollment}_{ij} = 1 \) if child \( i \) from Head Start center \( j \) is enrolled in Head Start and 0 otherwise; \( \text{Center}_{mj} \) represents a set of site dummies – each dummy equals 1 if Head Start center \( j \) is Head Start center \( m \) and 0 if not; \( \text{RA}_{ij} = 1 \) if child \( i \) from Head Start center \( j \) was randomly assigned to the program and 0 if not; \( \text{Parenting}_{ij} \) represents each parenting outcome; \( X_{kij} \) represents each baseline covariate; \( \hat{\text{Enrollment}}_{ij} \) is the predicted value of Head Start enrollment (based on the estimated parameters for Eq. 4) for child \( i \) from Head Start center \( j \); \( e_{1ij} \) is a random error that is independently distributed across individuals with a mean of zero and a variance of \( \sigma_{T}^2 \) and \( \sigma_{C}^2 \) for treatment and control group members, respectively; and \( e_{2ij} \) is a random error that is independently distributed across individuals with a mean of zero and a variance of \( \sigma_{T}^2 \) and \( \sigma_{C}^2 \) for treatment and control group members, respectively. Notably, the parameter \( \gamma_m \) in the first-stage equation is the mean effect of random assignment to Head Start on the probability of enrolling in Head Start for children from center \( m \), and the parameter \( \delta_m \) in the second-stage equation is the mean effect of Head Start enrollment on the outcome for children from center \( m \).

Standard errors of the parameter estimates for Eq. 5 were estimated using an approach developed by Brachet (2007) and employed by Bloom and Weiland (2015). In this approach, the predicted residuals from Eq. 5 are regressed on all the independent variables from Eq. 5 while clustering the standard errors by site (for a proof of this method, see Brachet, 2007).
This procedure produces estimates of the mean Head Start enrollment effect for each Head Start center ($\hat{\delta}_j$) and its standard error ($\hat{SE}(\hat{\delta}_j)$). The square of the standard error is its estimated error variance ($\hat{V}_j$).

**Step 2.** The second step of this approach involved inputting the values of $\hat{\delta}_j$ and $\hat{V}_j$ into a random-effects meta-analysis in order to obtain an estimate of the cross-site variation in Head Start enrollment effects on parenting. To do so, I estimated the following model:

**Center-level model:**

$$\delta_j = \delta_0 + w_j$$  

[Eq. 6]

where $\delta_j$ is the mean effect of Head Start enrollment on each parenting outcome for children from center $j$, $\delta_0$ is the cross-site grand mean effect of Head Start enrollment on the outcome, and $w_j$ is a random error that varies independently and identically across centers with a mean of zero and a variance of $\tau^2_{LATE}$. The statistical significance of $\delta_0$ was tested using the standard $z$ statistic from the random-effects meta-analysis, which was obtained using the MVMETA command in Stata 15.0. The statistical significance of $\tau^2_{ITT}$ was tested using the conventional Q-statistic for random-effects meta-analysis (Bloom et al., 2017; Bloom & Weiland, 2015; Hedges & Olkin, 1985; Weiss et al., 2017).

**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 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 pre-existing differences between the groups.

**OLS versus FIRC Approaches**

As a reminder, the FIRC approach employed in the present study more precisely estimates true treatment effect variation than the more “traditional” OLS approach, which over-estimates treatment effect heterogeneity. For a visual representation of the difference in these approaches as applied to the present study, see Figure 2. While results from these two approaches tell similar stories with regard to the average treatment effect (represented by the vertical solid line), Figure 2 demonstrates that the distribution of effects estimated using the FIRC approach is “tighter” around the average treatment effect compared to the OLS approach. The wider distribution shown in the OLS models demonstrates that this approach over-estimates treatment effect heterogeneity, while the FIRC approach more accurately estimates true treatment effects and thus also more accurately estimates treatment effect heterogeneity.

**ITT Results**

For all models (ITT and LATE), effect sizes for continuous outcomes (literacy activities, math activities, number of spanks) 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 (Bloom & Weiland, 2015; Weiss et al., 2017). 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.

**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%) on average than did parents in the control group. Head Start parents also increased their literacy activities (by .157 standard deviation units) and their math activities (by .179 standard deviation units) 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 (see Figure 1 for a visual representation of this variation). 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 95% of centers would have an effect size for literacy outcomes between -0.28 and 0.59. 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 95% of centers would have an effect size for math activities between -0.21 and 0.57—again, a wide range of effects. Effects for the probability of reading daily did not vary significantly across centers.

**Harsh 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 significantly – though modestly – for the frequency with which parents spanked children (Figure 1). Specifically, the significant standard deviation of 0.019 around the average effect of -0.001 indicates that if effect sizes were normally distributed across centers, then 95% of centers would have effect sizes between -0.04 and 0.04 (-4% and 4%) for spanking frequency. This is not a very wide range. Together, results suggest that Head Start random assignment effects for harsh discipline are small and either do not vary or 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 only. 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) but smaller standard deviations around those effects in the LATE models compared to ITT models. Specifically, the average effect for literacy activities was 0.157 with a significant standard deviation of 0.221, indicating that if normally distributed, 95% of centers’ effect sizes would be between -0.15 and 0.54. The average effect for math activities was 0.236 with a significant standard deviation of 0.176, indicating that if normally distributed, 95% of centers’ effect sizes would be between -0.11 and 0.58. Larger average effects in LATE models likely reflect the fact that these were families who were actually participating in Head Start and thus benefitting from program services, whereas ITT effects are
likely diluted by non-compliers. Along the same lines, smaller standard deviations around those average effects in the LATE models compared to ITT models may be indicative of more precise site-level estimates attained by eliminating uncertainty introduced by non-compliers in LATE estimates.

**Harsh discipline.** Whereas random assignment to Head Start did not impact parents’ harsh discipline, enrollment in Head Start did reduce the likelihood of parents spanking their children (by 3.7%). Enrollment did not impact parents’ 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 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 Head Start’s average effect 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 Head Start random assignment and participation 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 also varied significantly for two of these three outcomes—literacy and math activities. Effects were much weaker for Head Start’s effects on parents’ harsh discipline behavior. Only enrollment—but not random assignment—had a significant effect on decreasing parents’ likelihood of spanking, and this effect was small in magnitude. Effects on parents’ likelihood of spanking did not vary significantly across Head
Start programs. Finally, neither random assignment nor enrollment impacted parents’ frequency of spanking, though the effect of random assignment varied slightly across programs.

Consistent with prior reports (Gelber & Isen, 2013; Puma et al., 2010), results from the present study indicate that Head Start is effective at promoting parents’ cognitively stimulating interactions with children. This study extends 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. 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 in their ability to do so. 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 their young children. Perhaps it is easier or more straightforward to convey the importance of daily reading with children than are messages about multiple additional literacy and math activities, thus Head Start is more consistently good at promoting this outcome among its parents.

Head Start’s impacts on parents’ harsh discipline behaviors were minimal. Only enrollment, and not random assignment, significantly predicted the average treatment effect on parents’ likelihood of spanking. At first glance, this finding seems to contradict prior studies finding that Head Start reduces parents’ use of harsh 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 seven 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 average effects on spanking were smaller than in the original impact report and not statistically significant. Furthermore, it might be the case that Head Start is only effective at reducing parents’ spanking for children enrolled in—and not randomly assigned to—Head Start. Non-experimental studies estimate the association between Head Start enrollment (rather than random assignment) and parents’ behavior while controlling for family characteristics. Specifically, studies examining parents’ harsh discipline have found that Head Start parents reduced their spanking likelihood more than parents whose children were in parental care by 16% (Magnuson & Waldfogel, 2005) and compared to children in other center-based arrangements by 38-46% (Zhai et al., 2013). Although non-experimental, these findings support the possibility that Head Start effects on spanking may be specific to parents enrolled in Head Start. On the other hand, estimates from these studies were much larger than those found in the present study, in which I found that parents whose children enrolled in Head Start were about 4% less likely to spank children than parents who complied with assignment to the treatment group. Thus, although significant and in line with prior studies considering Head Start enrollment, causal estimates suggest that Head Start reduced parents’ spanking only slightly.

Results from the present study also indicate that Head Start effects on parenting 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. Head Start was thus relatively consistently ineffective at lowering parents’ spanking use.

When taken together, results demonstrate that Head Start is effective on average at promoting parents’ cognitive stimulation but that these effects on cognitive stimulation still vary widely, while Head Start is consistently ineffective—or at least consistently minimally effective—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 widely 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 non-physical discipline, or might make parents less susceptible to these suggestions even if they are made. 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, Goodman, Miller-Perrin, Holden, & Kazdin, 2018). Indeed, low-income families both endorse spanking and use spanking as a disciplinary strategy at higher rates than do their higher-income counterparts (Hines, Ryan, Kalil, & Ziol-Guest, 2019; Ryan et al., 2016).
While Head Start was more effective at improving parents’ cognitively stimulating interactions with children than their disciplinary interactions with them, Head Start was still not consistently effective at promoting cognitive stimulation. Indeed, some programs were actually less effective at promoting parents’ literacy and math activities than their local alternatives, meaning that these parents likely would have improved more had they not been randomly assigned to or attended Head Start. Thus, while Head Start may be prioritizing parents’ enriching activities with their children, the fact that the program is not universally effective indicates that there is still work to be done in this domain so that Head Start can be maximally effective.

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 mothers in the control condition were free to choose 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 a somewhat fuzzy comparison group. This means that it is not possible to produce causal estimates of Head Start compared to other ECE arrangement types, though prior work demonstrates that estimates may vary depending on which alternative care arrangement Head Start is being compared to (Morris et al., 2018; Zhai et al., 2013). Finally, while the HSIS is the only multi-site RCT of Head Start programs and thus is the ideal dataset for 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 pre-k programs, the Head Start treatment contrast has changed significantly since 2002. More specifically, the control group in 2002 may have had access to a very different set of options when compared to children today and thus findings may or may not be generalizable to Head Start families today. Head Start still leads the ECE landscape in terms of its attention to parents, though, so the contrast between outcomes for parents whose kids 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 all parents, including low-income parents, doing significantly more literacy and math activities with children in 2012 than in 1998 (Kalil, Ziol-Guest, Ryan, & Markowitz, 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 low-income parents are already doing all they can and do not have much room to grow, but may also be more effective at changing parenting if low-income parents today are more motivated to be involved in their children’s learning. Indeed, despite increases in literacy and math activities with children, low-income parents have not yet caught up to the levels of these parenting behaviors observed in higher-income families (Kalil et al., 2016).

Despite these limitations, this study makes important contributions to the literature by demonstrating that there is wide variation in program effects on parents’ cognitively stimulating interactions with their children and only minimal variation in program effects on parents’ harsh discipline. Knowing that programs vary in their effectiveness on at least some parenting outcomes begs the question—what factors predict which centers are more versus less effective?
The next two chapters of this dissertation explore multiple center characteristics (chapter 3) and family characteristics (chapter 4) that could uncover why programs differ in their effectiveness at improving parenting practices.
Table 1
Descriptive statistics and baseline balance of the analytic sample

<table>
<thead>
<tr>
<th>Parenting behaviors at baseline (Fall 2002)</th>
<th>Combined mean/% (N=3,527)</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>Control group mean/% (N=1,357)</th>
<th>Treatment group mean/% (N=2,170)</th>
<th>Difference</th>
<th>p-value of difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive stimulation</td>
<td>1.162</td>
<td>0.374</td>
<td>0</td>
<td>2</td>
<td>1.129</td>
<td>1.183</td>
<td>0.054</td>
<td>0.000 ***</td>
</tr>
<tr>
<td>Child is read to daily (%)</td>
<td>0.345</td>
<td>0.475</td>
<td>0</td>
<td>1</td>
<td>0.331</td>
<td>0.354</td>
<td>0.023</td>
<td>0.160</td>
</tr>
<tr>
<td>Child spanked in previous week (%)</td>
<td>0.455</td>
<td>0.498</td>
<td>0</td>
<td>1</td>
<td>0.478</td>
<td>0.441</td>
<td>-0.037</td>
<td>0.032 *</td>
</tr>
<tr>
<td>Number of spanks</td>
<td>0.892</td>
<td>1.442</td>
<td>0</td>
<td>15</td>
<td>0.976</td>
<td>0.840</td>
<td>-0.136</td>
<td>0.009 **</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Child characteristics</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>0.595</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child was in 3-year-old cohort (%)</td>
<td>0.553</td>
<td>0.497</td>
<td>0</td>
<td>1</td>
<td>0.552</td>
<td>0.554</td>
<td>0.002</td>
<td>0.909</td>
</tr>
<tr>
<td>Female (%)</td>
<td>0.500</td>
<td>0.500</td>
<td>0</td>
<td>1</td>
<td>0.494</td>
<td>0.504</td>
<td>0.009</td>
<td>0.595</td>
</tr>
<tr>
<td>Black (%)</td>
<td>0.302</td>
<td>0.459</td>
<td>0</td>
<td>1</td>
<td>0.295</td>
<td>0.306</td>
<td>0.011</td>
<td>0.479</td>
</tr>
<tr>
<td>Hispanic (%)</td>
<td>0.371</td>
<td>0.483</td>
<td>0</td>
<td>1</td>
<td>0.371</td>
<td>0.371</td>
<td>0.000</td>
<td>0.999</td>
</tr>
<tr>
<td>White (%)</td>
<td>0.327</td>
<td>0.469</td>
<td>0</td>
<td>1</td>
<td>0.334</td>
<td>0.323</td>
<td>-0.011</td>
<td>0.490</td>
</tr>
<tr>
<td>Home language is Spanish (%)</td>
<td>0.295</td>
<td>0.456</td>
<td>0</td>
<td>1</td>
<td>0.293</td>
<td>0.296</td>
<td>0.003</td>
<td>0.835</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Family characteristics</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>0.560</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maternal age (years)</td>
<td>29.170</td>
<td>7.448</td>
<td>16</td>
<td>78</td>
<td>28.977</td>
<td>29.291</td>
<td>0.314</td>
<td>0.218</td>
</tr>
<tr>
<td>Maternal education (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Less than HS</td>
<td>0.380</td>
<td>0.485</td>
<td>0</td>
<td>1</td>
<td>0.398</td>
<td>0.368</td>
<td>-0.030</td>
<td>0.078 +</td>
</tr>
<tr>
<td>HS</td>
<td>0.332</td>
<td>0.471</td>
<td>0</td>
<td>1</td>
<td>0.326</td>
<td>0.336</td>
<td>0.009</td>
<td>0.560</td>
</tr>
<tr>
<td>More than HS</td>
<td>0.288</td>
<td>0.453</td>
<td>0</td>
<td>1</td>
<td>0.276</td>
<td>0.296</td>
<td>0.020</td>
<td>0.194</td>
</tr>
<tr>
<td>Maternal marital status (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.557</td>
</tr>
<tr>
<td>Single</td>
<td>0.395</td>
<td>0.489</td>
<td>0</td>
<td>1</td>
<td>0.393</td>
<td>0.396</td>
<td>0.004</td>
<td>0.835</td>
</tr>
<tr>
<td>Married</td>
<td>0.448</td>
<td>0.497</td>
<td>0</td>
<td>1</td>
<td>0.455</td>
<td>0.444</td>
<td>-0.011</td>
<td>0.527</td>
</tr>
<tr>
<td>Previously married</td>
<td>0.157</td>
<td>0.364</td>
<td>0</td>
<td>1</td>
<td>0.153</td>
<td>0.160</td>
<td>0.007</td>
<td>0.557</td>
</tr>
<tr>
<td>Mother is a teenager (%)</td>
<td>0.164</td>
<td>0.370</td>
<td>0</td>
<td>1</td>
<td>0.173</td>
<td>0.158</td>
<td>-0.015</td>
<td>0.242</td>
</tr>
<tr>
<td>Mother is a recent immigrant (%)</td>
<td>0.186</td>
<td>0.389</td>
<td>0</td>
<td>1</td>
<td>0.191</td>
<td>0.183</td>
<td>-0.008</td>
<td>0.558</td>
</tr>
</tbody>
</table>
### Assessment characteristics

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Child age at spring testing (years)</td>
<td>4.528</td>
<td>0.567</td>
<td>3.250</td>
<td>6.288</td>
<td>4.535</td>
<td>4.524</td>
<td>-0.011</td>
<td>0.571</td>
</tr>
<tr>
<td>Child was tested in Spanish (%)</td>
<td>0.249</td>
<td>0.433</td>
<td>0</td>
<td>1</td>
<td>0.243</td>
<td>0.253</td>
<td>0.010</td>
<td>0.511</td>
</tr>
</tbody>
</table>
| Time between baseline and spring parent interview (weeks) | 33.442 | 3.309  | 26     | 48     | 33.808 | 33.213 | -0.594 | 0.000  ***

**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

<table>
<thead>
<tr>
<th></th>
<th>Child is read to daily (y/n)</th>
<th>Literacy Activities</th>
<th>Math activities</th>
<th>Child spanked in past week (y/n)</th>
<th>Number of spanks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$p$</td>
<td>$p$</td>
<td>$p$</td>
<td>$p$</td>
<td></td>
</tr>
<tr>
<td><strong>Effects of assignment to Head Start (ITT)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average effect</td>
<td>0.044</td>
<td>0.005</td>
<td>0.157</td>
<td>0.000</td>
<td>0.179</td>
</tr>
<tr>
<td>Standard deviation of effect</td>
<td>0.059</td>
<td>0.152</td>
<td>0.221</td>
<td>0.000</td>
<td>0.201</td>
</tr>
<tr>
<td>Range of effects</td>
<td>(-0.07, 0.16)</td>
<td>(-0.28, 0.59)</td>
<td>(-0.21, 0.57)</td>
<td>(-0.09, 0.05)</td>
<td>(-0.04, 0.04)</td>
</tr>
<tr>
<td><strong>Effects of participation in Head Start (LATE)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average effect</td>
<td>0.052</td>
<td>0.008</td>
<td>0.199</td>
<td>0.000</td>
<td>0.236</td>
</tr>
<tr>
<td>Standard deviation of effect</td>
<td>0.077</td>
<td>0.195</td>
<td>0.174</td>
<td>0.003</td>
<td>0.176</td>
</tr>
<tr>
<td>Range of effects</td>
<td>(-0.10, 0.20)</td>
<td>(-0.14, 0.54)</td>
<td>(-0.11, 0.58)</td>
<td>(-0.04, 0.04)</td>
<td>(-0.05, 0.04)</td>
</tr>
</tbody>
</table>

| N children                           | 3,516                        | 3,416              | 3,462          | 3,525                           | 3,512           |
| N centers                            | 297                          | 296                | 297            | 297                             | 296             |

Notes: 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 HSIS 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, number of spanks) 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.

1 Assuming normality, 95% of cases would fall in this range.

***p<.001, **p<.01, *p<.05, +p<.10
Figure 1. Cross-site distributions of Head Start ITT effects by outcome measure.
Figure 2. Cross-site distributions of Head Start ITT effects by outcome measure based on the OLS with treatment-by-center interactions approach (dashed line) and the fixed intercept, random slopes approach (solid line).
CHAPTER III: CENTER CHARACTERISTICS AS SOURCES OF VARIATION IN HEAD START EFFECTS ON PARENTING OUTCOMES

The analyses presented in chapter two demonstrated that there was meaningful variation in Head Start treatment effects on parents’ cognitively stimulating interactions with their children and, to a smaller degree, the frequency with which parents spank their children. This variation indicates that some Head Start centers are more effective than others at improving parenting outcomes relative to their local alternatives, including parent care. A natural next line of inquiry when studying treatment effect variation is: what factors predict whether a program is more effective than others?

In their conceptual framework for studying the sources of variation in program effects, Weiss and colleagues (2014) posit three factors that might moderate the causal relationship between a particular intervention’s treatment and the program’s treatment effects: characteristics of the organization; characteristics of the individuals participating in the intervention; and the treatment contrast, which is the difference between what the treatment and control group of an intervention experience. The present study examines the first of these sources of variation in Head Start effects on parenting outcomes: characteristics of Head Start centers themselves.

When thinking about the types of things Head Start centers do to help facilitate positive parenting behaviors, two features of Head Start programs are particularly relevant. The first is a center’s parent outreach efforts, which are efforts made by the center to promote parents’ active engagement in their children’s development. The second is the center’s average classroom quality, which refers to whether the center provides a safe environment, adequate learning resources, and warm and sensitive teacher-child interactions.

The present study expands on analyses presented in Chapter 2 by asking whether these two theoretically important features of Head Start centers—parent outreach and classroom
quality—moderate the treatment effects on parents’ cognitively stimulating and disciplinary interactions with their children in the HSIS. That is to say, in this study I test whether Head Start treatment effects on parents’ outcomes are different for centers high versus low in both parent outreach and center quality. Results from the present study will help to identify aspects of Head Start programs most strongly associated with improvements in parenting. Findings thus have important implications for helping Head Start decision makers understand where efforts to strengthen parent outcomes might be best directed.

Parent Outreach in Head Start

Engaging families in their children’s healthy development has been a goal of Head Start since its inception (Zigler & Styfco, 2010; Zigler & Valentine, 1979), and Head Start remains committed to this goal today. The current Head Start Performance Standards state that Head Start programs “must promote shared responsibility with parents for children’s early learning and development, and implement family engagement strategies that are designed to foster parental confidence and skills in promoting children’s learning and development,” (Administration for Children and Families, 2016). These standards go on to specify that family engagement strategies must include offering activities that support parent-child relationships and child development.

Parent outreach activities occur both inside and outside of Head Start centers. Some of the opportunities that programs must offer parents and family members inside centers include regular communication between teachers and parents to ensure parents are well-informed about their children’s routines, activities, and behavior, as well as regular (as needed, but no less than two times per program year) teacher-parent conferences to enhance the knowledge and understanding of both staff and parents of the child’s education and developmental progress and
activities in the program (Administration for Children and Families, 2018). Outside of centers, Head Start must offer regular home visits (at least two per year, including one before the program year begins, if feasible) to engage parents in their child’s learning and development.

As a way to guide Head Start grantees with implementing the parent engagement requirements of the Head Start performance standards, in 2016 the Administration for Children and Families introduced the Parent, Family, and Community Engagement (PFCE) Framework (Administration for Children and Families, 2018). The PFCE Framework stresses program-wide strategic planning; integration of services across Head Start and community agencies; joint responsibility for children’s healthy development between Head Start staff (including center directors, teachers, family service workers, and home visitors) and parents and other family members; and the inclusion of all of these parties in promoting positive child and family outcomes (Administration for Children and Families, 2018). This framework simultaneously makes clear that all Head Start staff are required to make efforts to engage parents and other family members in children’s learning while also leaving individual grantees with freedom to decide what this partnership will look like and how it will work best at their center, given their specific family and community needs. There is therefore variation in how individual centers approach family engagement (Aikens et al., 2017) and thus also variation in the quality and quantity of outreach activities that individual centers perform. Knowing the extent to which outreach activities promote at-home parental involvement in children’s learning is thus of great interest to Head Start decision makers.

The Link Between Parent Outreach Activities and Parenting Behaviors

Theoretically, centers that make greater efforts to engage parents in their children’s learning will be more successful at facilitating parents’ positive engagement with children in the
home. These efforts are often referred to as “parent outreach” in order to distinguish efforts that the centers themselves make to engage parents from parent-motivated engagement in Head Start, such as parents’ time volunteering in the classroom or on the Head Start policy council, helping plan school events, or chaperoning field trips (for which there is a separate literature that I will not review here; see Ansari & Gershoff, 2016 for more details). Any observed effects of parent outreach efforts cannot be attributed to selection factors that make some parents more likely than others to be engaged in Head Start because parent outreach is initiated by the center, not by the parent.

Parent outreach efforts occur both inside and outside of centers themselves. Parent outreach that occurs inside centers includes activities like workshops, formal meetings between teachers and parents, and even informal time that parents and teachers spend together during pick-up or drop-off (Aikens, 2017). Parent outreach that occurs outside the center comprises home visiting services, which deliver a range of services to families in their homes. The idea behind home visiting is grounded in theoretical frameworks that presume that parenting behavior and parent-child interaction is shaped by parents’ psychological resources and contextual sources of stress and support (Belsky, 1984; Bronfenbrenner, 1989; Bronfenbrenner & Morris, 2006). As a result, home-visiting programs not only provide direct modeling and instruction about optional parenting practices but also provide services to ameliorate stressors that can undermine parenting, including maternal mental health problems, low social support, and financial strain (Ryan & Padilla, 2018).

During all of these times, parents are exposed to the types of direct instruction and modeling that were theorized to be the mechanism by which Head Start impacts parenting outcomes in Chapter 1. For example, parents can learn explicitly from providers about effective
or otherwise positive parenting behaviors – or about child development – in ways that could enhance their parenting (Brooks-Gunn et al., 2000; Howard & Brooks-Gunn, 2009; Lee et al., 2014; Rafferty et al., 2010). Depending on the provider’s credentials or knowledge, he or she may suggest ideas (solicited or not) for activities parents could do with children to further their learning, which the parent might then implement with the child at home (McCartney et al., 2007). With regard to discipline, a provider might be able to suggest strategies to effectively address misbehavior through non-physical means, thus reducing the frequency with which parents use harsh discipline. Providers might also give parents information about children’s development, which could in turn improve parenting practices (Bartlett, Guzman, & Ramos-Olazagasti, 2018). For example, a parent who is unaware that young children benefit from early language and math exposure might learn from her child’s provider that children can make gains in these domains at very early ages through activities and experiences and might modify her behavior accordingly.

The previous literature on the link between parent outreach activities that occur in centers or in schools and parents’ outcomes is sparse. To my knowledge, no prior study has examined the specific link between ECE centers’ outreach aimed at increasing parent involvement in the home and parenting outcomes. However, there are some studies on related constructs that lay the groundwork for thinking about whether parent outreach predicts variation in Head Start treatment effects. For example, in a study examining the extent to which schools’ outreach to families was associated with family involvement in the home and at school, Galindo and Sheldon (2012) found that schools’ parent outreach efforts were associated with parental involvement with their kindergarteners at school, but not at home. Authors caution, however, that their measure of parent outreach focused heavily on activities aimed at increasing parents’
participation in events at school (e.g. inviting parents to school performances and classroom programs, fundraising activities, parent-teacher-student organization meetings, teacher-parent conferences, and workshops for teachers that focus on parent involvement) and thus perhaps a different measure more focused on encouraging at-home involvement would have predicted increased parent involvement at home. More specific to Head Start, in a set of studies using two cohorts of the Head Start Family and Child Experiences Survey (FACES), authors conceptualized family outreach via the primary outreach-related goals that Head Start center directors reported setting (e.g. informing parents about child development, supporting family well-being, and encouraging participation in Head Start decision making). Results showed that when centers listed promoting parent-child book reading as one of their top three primary outreach goals, parents in the FACES: 2003 cohort increased their parental cognitive stimulation with children more over the course of a year of Head Start than parents at centers that listed other priorities (Hindman & Morrison, 2011). If goal setting is indicative of the activities that centers actually do, then this study suggests that centers that do more outreach activities may be more effective at enhancing parents’ cognitive stimulation than centers that do fewer such activities. Centers’ goals did not predict parent involvement in the home in the FACES: 2000 cohort, however (Hindman, Miller, Froyen, & Skibbe, 2012), indicating that either centers’ goals are not indicative of their actual outreach activities, or that the link between outreach and parenting outcomes is tenuous.

Studies examining the link between parent outreach and children’s outcomes also shed light on whether outreach is able to promote parenting outcomes. For example, controlling for factors that might bias the association between outreach and child outcomes (e.g. poverty, prior math performance), one study found that elementary schools that implemented more math-
related parent outreach activities (e.g. conducting workshops to help parents understand how to work with children at home to strengthen math skills, conducting parent-teacher conferences to discuss how to help children improve in math) had higher levels of student math achievement after one year (Sheldon, Epstein, & Galindo, 2010). Similarly, a meta-analysis that examined the associations between the addition of parenting education services to ECE programs and children’s outcomes found that ECE centers that provide more opportunities for parents to learn through modeling or opportunities to practice stimulating behaviors and those that provide one or more home visits per month and were more effective at improving children’s short-term reading and math outcomes than programs that provided fewer opportunities to observe and practice parenting skills or fewer home visits (Grindal et al., 2016). Although none of these studies specifically tested the mechanism by which parent outreach predicted improved child outcomes, presumably they did so by encouraging parent involvement in the home. Thus, it is likely that Head Start centers that do more parent outreach are more effective at facilitating positive parenting outcomes than programs that do less outreach.

Home visiting may be a particularly promising avenue by which Head Start impacts parenting, given that home visiting services explicitly target parenting outcomes. Home visiting often serves as a stand-alone support or intervention for new parents. Well-established home visiting programs like Healthy Families America and Nurse-Family Partnership have shown success in enhancing parents’ responsivity during parent-child interactions, promoting cognitive stimulation in the home, and reducing harsh parenting as well as child abuse and neglect (Duggan et al., 2007; Easterbrooks et al., 2013; Eisenberg et al., 1995; Kitzman et al., 2000; Olds et al., 1997; Rodriguez, Dumont, Mitchell-Herzfeld, Walden, & Greene, 2010).
Home visiting is also sometimes offered through children’s ECE services. The best example of home visiting services offered through or in conjunction with ECE services comes from Early Head Start, which is a precursor to Head Start for low-income pregnant women and families with children under age three. Early Head Start operates via three different program approaches: traditional center-based programs, which provide services mainly in center-based care along with a minimum of 2 home visits per year; home-based programs, which provide services to families mainly through weekly or biweekly home visits and at least 2 parent-child group socialization activities per month; and mixed-approach programs, which provide services through a combination of center-based care and home-based approaches. In a large national evaluation of Early Head Start programs across 17 sites, the mixed-approach programs were found to be most effective at improving parenting outcomes. Specifically, parents in mixed-approach programs were found to be more likely to read to children daily, less likely to spank children, and more supportive and less detached during parent-child interactions at age three compared to parents in the control condition (Love et al., 2005). There were fewer significant effects for parents in home-based programs—those parents were found to be more supportive during parent-child interactions than parents in the control condition. No statistically significant impacts on parenting outcomes were found among families in the center-based only programs. This evaluation provides strong support for the potential for home visiting in conjunction with children’s ECE services to promote positive parenting behaviors in the domains of both cognitive stimulation and socioemotional interactions, including discipline.

Specific to Head Start itself, there is some evidence suggesting that home visiting services offered through Head Start can impact parenting outcomes. For example, a study examining the association between Head Start parents’ engagement in home visiting and
parenting outcomes found that parents who participated in more home visiting sessions had more positive interactions with children than parents who did not—but only if the home visiting sessions were delivered by a facilitator with whom the parents were familiar (Stormshak, Kaminski, & Goodman, 2002). This study did not randomly assign participants to more versus less home visits, however, and thus findings should not be interpreted as causal. However, an RCT of an intervention (Research-based Developmentally Informed-Parent program, or REDI-P) implemented through Head Start and designed to aid parents during their children’s transition to kindergarten demonstrated positive outcomes (Bierman, Welsh, Heinrichs, Nix, & Mathis, 2015). REDI-P delivered approximately 10 home visits to parents during Head Start and then six “booster” sessions once children had entered kindergarten. Parents randomly assigned to receive the REDI-P intervention reported reading more interactively with children and having longer and more frequent conversations with them than parents in the control condition (Bierman et al., 2015). However, observational ratings of parents’ support for learning did not differ for treatment and control groups. These findings lend some support for the idea that Head Start centers that provide more home visits may be more effective at improving parenting outcomes than those that provide fewer home visits.

Further support comes from an evaluation of variation in Head Start treatment effects on children’s outcomes using the HSIS. Importantly, although this study examined child and not parent outcomes, this is the only study to explicitly test the role of home visiting in explaining variation in Head Start treatment effects (as opposed to studying the treatment effect for a single center) and is thus particularly relevant to the present investigation. In this study, Walters (2015) found that centers that provided four or more home visits per year were more effective at improving children’s non-cognitive skills (by .11 standard deviations) than centers providing
fewer home visits. Although this study did not directly test the mechanism by which home visiting improved child outcomes, it is reasonable to presume that it did so through its effect on positive parenting behaviors. Thus, it may be that centers that offer more home visits are also more effective at promoting parents’ positive outcomes than centers that offer fewer home visits.

**Classroom Quality in Head Start**

Another center characteristic theoretically important for determining Head Start centers’ impacts on parenting is classroom quality. High-quality ECE settings are characterized by warm and responsive interactions between children and providers that support children’s learning and self-regulation and foster higher-order thinking and social skills (Mashburn et al., 2008; Yoshikawa et al., 2013). Quality settings also provide access to a range of early learning materials such as books, blocks, and sensory experiences.

Classroom quality is made up of two components: process quality and structural quality. Process quality features are those pertaining to the quality of teacher-child interactions, which are ideally positive and stimulating (Yoshikawa et al., 2013). Structural features of quality are those that can be changed by structuring the setting differently or putting different requirements for staff in place, and include features like group size, ratio, and teacher qualifications. Structural characteristics of quality enable higher quality process features, which are the most important contributors to children’s skills (Yoshikawa et al., 2013). Specific dimensions of process quality include emotional support (provider sensitivity and responsiveness to children’s needs) instructional support (provider promotion of concept development and scaffolding of skill development), and classroom organization (more effective management of peer interactions and setting of behavioral norms).
The evidence on the average quality of Head Start programs compared to other ECE programs is mixed (Bassok et al., 2016; Dowsett, Huston, Imes, & Gennetian, 2008; Feller et al., 2016; Johnson, Ryan, & Brooks-Gunn, 2012), but Head Start is generally found to be among the higher quality programs that low-income families use in terms of classroom organization, teacher qualifications, and teachers’ cognitive stimulation (Bassok et al., 2016; Dowsett et al., 2008; Johnson et al., 2012). Within Head Start programs themselves, there is notable variation in classroom quality (Puma et al., 2010). This variation tends to be small in magnitude, however, due to most programs scoring in the “minimal-good” range of quality on the ECERS, which is a global measure of quality (Alamillo et al., 2018; Moiduddin, Aikens, Tarullo, West, & Xue, 2012; Puma et al., 2010) and most Head Start classrooms scoring in the low range on the CLASS, which measures process quality (Alamillo et al., 2018; Moiduddin et al., 2012).

The Link Between Classroom Quality and Parenting Behaviors

With regard to predicting parenting outcomes, it is reasonable to hypothesize that higher-quality centers—particularly those high in process quality but also those high in structural quality—are more likely to enhance positive parenting practices than lower quality ones, given their greater capacity to provide direct instruction and modeling to parents. For example, a parent can only learn more stimulating and sensitive behaviors from providers if the providers’ behavior with children is itself stimulating and sensitive, and centers that have smaller or more regulated classrooms may have more of the organizational resources necessary to effectively reach, engage, and communicate with parents. It is also more likely that providers in high-quality settings are willing and able to provide childrearing guidance and advice and to impart knowledge of child development in such a way as to elicit more positive behavior on the part of the parents than providers in lower-quality settings.
There has been very little research on the effect of ECE quality on parenting behaviors, but a few studies using the National Institute of Child Health and Human Development (NICHD) Study of Early Child Care and Youth Development (SECCYD) can tell us about this association. One set of these studies found that higher quality child care predicted greater maternal sensitivity across children’s first three years of life (NICHD ECCRN, 1999), and that this positive association extended through first grade, but only for children who had experienced relatively few hours of ECE (NICHD ECCRN & Duncan, 2003). In another, McCartney and colleagues (2007) found that higher quality child care was associated with improvements in the quality of the home environment (which included maternal responsiveness, maternal cognitive stimulation, as well as availability of learning materials in the home) from 6 to 36 months. This improvement was most pronounced for poorer families, suggesting that this relationship may also exist in Head Start settings.

It is not currently known whether Head Start quality is associated with more positive parenting outcomes, but if Head Start center quality does predict parenting outcomes, then higher quality centers should be more effective at promoting positive parenting behaviors than lower quality centers. However, given prior reports suggesting that the variation in Head Start quality may be relatively small (Alamillo et al., 2018; Puma et al., 2010; Moiduddin et al., 2012), it may not be possible to detect variation in parenting outcomes as a result of classroom quality in the HSIS.

**The Present Study**

The present study explores whether two theoretically important characteristics of Head Start for improving parenting behaviors—parent outreach efforts and classroom quality—might explain why some Head Start centers are more effective at improving parenting outcomes than
others. Specifically, I ask whether centers that offer more parent outreach activities, centers that offer more home visits, and centers that are higher quality are more effective than centers with less parent outreach, centers that offer fewer home visits, and centers that are lower quality. I predict that centers whose providers offer more outreach activities will be more effective than those that offer fewer outreach activities, that centers that require more home visits will be more effective than centers that offer fewer home visits, and that centers with higher average classroom quality will be more effective than centers with lower average classroom quality.

It is important to note that although only effects on parents’ math and literacy activities—and to a smaller and less consistent degree, spanking frequency—were shown in Chapter 2 to vary significantly across centers, I examine whether outreach or quality predict treatment effect variation for the full set of parenting outcomes included in Chapter 2. This is because some prior work has shown that even in the absence of significant treatment effect variation overall, it may still be possible to predict treatment impact variation with theoretically-informed site-level characteristics (Bloom & Spybrook, 2017; Sabol et al., 2019). This is because even if the overall cross-site treatment effect variation is not large enough to reach statistical significance, the moderating effect of a single theoretically-informed variable on the treatment effect could be large enough to be significant. Thus, it is recommended to test predictors of treatment impact variation even if overall treatment impact variation is not shown to vary significantly (Sabol et al., 2019).

Methods

Data and Sample

Data again come from the HSIS. Sample restrictions were made in the same way as they were for Study 1, with the exception that samples were further restricted to cases with non-
missing data on the center characteristic under investigation. This process resulted in sample sizes that ranged from 3,251-3,465 children within 276-286 Head Start Centers.

**Measures**

All parenting measures and covariates in the present study were identical to those included in Chapter 2. The following center characteristics were also included in this study.

**Parent outreach.** Center directors reported on the number of activities that the center’s lead teacher and the center’s family service worker performed. Specifically, center directors reported on the item “Which of the following activities do [lead teachers/family service workers] perform in the center?” with the following 9 parent-directed (as opposed to child-directed) options: informing parents about the progress of their child; teaching parents about parenting/education/child development issues including activities to do with their children; conducting family assessments; providing guidance to families to help them meet their goals; providing referral to community services; providing informal counseling or addressing personal issues (e.g. marital stress/family relations); providing information/referral to parents about educational services; providing assistance with basic needs (e.g. food/housing/clothing/medical care); obtaining information from parents about their experiences with the center including suggestions for improvement. Separate outreach measures were created for the lead teachers and family service workers. Because center directors reported that nearly half of family service workers (49.8%) did all 9 outreach activities, centers were considered to have “high” family service worker parent outreach if the center director reported that the family service worker did all 9 of those outreach activities. In order to make the two outreach measures analogous, centers were also considered to have “high” lead teacher parent outreach if the center director reported that the lead teacher did all 9 of those outreach activities (which represented 22.7% of centers).
**Home visiting.** Center directors reported on the item “Are visits to children’s homes required of any center staff?” if the response was “yes,” center directors reported on the item “How many home visits are required per program year?” Following Walters (2015), this item is a dummy variable indicating whether the center provided 0-3 home visits or more than 3 home visits. Twenty-one percent of centers required 4 or more home visits.

**Classroom quality.** Direct observations of care settings’ quality were used for children in center-based and family day care home programs. Trained observers conducted observations in classrooms and centers attended by the sampled children. Observers spent four hours in each class to ensure observation of a major portion of the daily schedule and a variety of classroom and center activities. Center quality was measured in two ways. First, classroom observers assessed quality using the Early Childhood Environmental Rating Scale-Revised (ECERS-R) (Harms, Clifford, & Cryer, 1998). Classroom observations conducted by study staff gave the classroom a rating ranging from a low of one (“inadequate”) to a high of seven (“excellent”) on 37 items covering six subscales: (1) adequacy of space and furnishings; (2) personal care routines; (3) language and reasoning, including materials available and activities used; (4) range of activities that are used and available; (5) interactions, including both staff-child and child-child interactions; and (6) program structure, including the use of a daily schedule. This variable was created by taking the total ECERS-R score for each Head Start classroom and averaging scores across classrooms within each center. As in the HSIS final report (Puma et al., 2010), this variable was then dichotomized to represent low quality (score below 4) or high quality (score of 5 or above). Sixty-five percent of centers scored “high” in classroom quality according to this cutoff.
Classroom observers also completed the Arnett Caregiver Interaction Scale (CIS) (Arnett, 1989), with ratings ranging from zero (“not at all”) to three (“very much”) on traits and interactions: (1) greater teacher sensitivity, (2) responsiveness, (3) encouragement of children’s independence, and (4) lower levels of punitiveness and detachment. Items indicating punitiveness and detachment were reverse coded to provide a measure of more positive teacher/child relationships. This variable was created by taking the total Arnett CIS score for each Head Start classroom and averaging scores across classrooms within each center. The variable was then dichotomized to represent low quality (score in the bottom 50%, which is a score below 2.59) or high quality (score in the top 50%, which is a score of 2.59 or above).

**Analytic Strategy**

To explore whether parent outreach or classroom quality moderate the Head Start treatment effect on each parenting outcome (i.e. whether they predict variation in treatment effects), I build on Equation 3 from the intent-to-treat (ITT) model presented in Chapter 2 by adding each center-level characteristic individually at level 2 as a predictor of the level 1 slope for treatment status. This creates a cross-level interaction term. Specifically, I estimate the following model:

**Level 1: Parents**

\[ Parenting_{T2ij} = \alpha_j + \beta_j RA + \lambda_i Parenting_{T1} + \sum_{k=1}^{K} \pi_k X_{kij} + \varepsilon_{ij} \]  

[Eq. 1]

**Level 2: Head Start Centers**

\[ \alpha_j = \alpha \]  

[Eq. 2]

\[ \beta_j = \beta_0 + \pi CenterCharacteristic_j + r_j \]  

[Eq. 9]

where all parameters in Eq. 1 and 2 are defined in the same way as in chapter 2; \( \beta_j \) is a slope that varies randomly across centers and represents the average effect of random assignment for Head
\( \beta_0^* \) represents the grand mean Head Start ITT effect size for centers where the value of the center characteristic in the model is zero; \( \pi \) is a slope for the non-causal “effect” of each center characteristic (lead teacher outreach, family service outreach, number of home visits, ECERS-R, and Arnett) on the mean Head Start treatment effect. This estimate represents the rate of change in the mean Head Start ITT effect size per point increase in the center characteristic in the model.

In these models, \( \pi \) is the main parameter of interest because it indicates whether and how much each center characteristic contributes to variation in the treatment effect. The statistical significance of \( \pi \) was tested using the corresponding \( z \) statistic from the 2-level mixed model, which was obtained using the MIXED command in Stata 15.0.

Note that I examine moderation of the treatment effect by center characteristics using only ITT and not LATE approaches. This is because methods for estimating moderation in a Local Average Treatment Effect (LATE) framework have not yet been developed and it would be computationally demanding to enter additional interaction terms into the two-step process for estimating LATEs, given the already large number of random assignment-by-center and predicted enrollment-by-center interactions included in these models.

**Results**

**Descriptive Results**

Descriptive results for center characteristics included as moderators in the present study are shown in Table 3. As discussed in the measures section, more centers scored high in family service worker outreach than in lead teacher outreach. Specifically, 22.7% of centers scored high in outreach by the lead teacher, and about half (49.8%) of centers scored high in outreach by the family service worker. Twenty-one percent of centers required 4 or more home visits. See Figure
3 for a visual representation of the variation in outreach scores in their original units (before they were dichotomized) across centers.

Sixty-five percent of centers scored “high” in classroom quality as measured by the ECERS-R, and about half (51.4%) scored “high” in classroom quality as measured by the Arnett CIS. See Figure 4 for a visual representation of the variation in quality scores in their original units (before they were dichotomized) across centers.

Pairwise correlation coefficients between each of the center characteristics included in the present study are displayed in Table 4. The size of correlations between center characteristics ranged in magnitude. The highest correlation was between the two measures of quality—ECERS-R and Arnett CIS scores ($r(283) = .44$, $p < .001$). The measures of lead teacher and family service worker parent outreach were also moderately correlated ($r(275) = .21$, $p < .001$). Family service worker outreach was positively correlated with home visiting frequency ($r(278) = .13$, $p = .028$), and lead teacher outreach was negatively correlated with ECERS-R scores ($r(270) = -.11$, $p = .083$). Correlations between all other center characteristics were small and not significant.

**Parent Outreach Results**

Results from models testing whether each center characteristic moderated the effect of Head Start random assignment on each parenting outcome are presented in Table 5. The “Treatment” coefficients listed under each center characteristic represent the estimated intercept, $\hat{\beta}_0^*$ from Eq. 9. This coefficient is interpreted as the ITT effect size for centers with a value of zero on the moderator in the model—for example, centers with low lead teacher outreach. The “Treatment*Moderator” coefficients represent $\pi$ from Eq. 9, the rate of change in the mean Head Start ITT effect size per point increase in the moderator. Because the moderators are
dichotomous, they reflect the difference in the ITT effect for those centers “high” in each moderator compared to those that were “low”.

High lead teacher outreach did not moderate the treatment effect for any of the examined outcomes, as indicated by the non-significant Treatment*High Outreach interaction coefficients presented in Table 5. This indicates that lead teacher outreach was not associated with centers’ effectiveness at improving parenting outcomes.

There was a negative interaction between Head Start random assignment and high family service worker outreach for parents’ daily reading and parents’ math activities. These significant interactions indicate that centers with high family service worker outreach were less effective at promoting parents’ likelihood of reading to children daily and at promoting their engagement in math activities than centers with lower family service worker outreach (Figure 5). More specifically, in low family service worker outreach centers, parents randomly assigned to Head Start were 8% more likely to read to their children daily compared to parents in the control condition who would have attended low-outreach centers had they been assigned to Head Start. By contrast, Head Start parents in high family service worker outreach centers were no more or less likely to read to children daily than parents in the control condition—the effect of random assignment for Head Start parents in high family service worker outreach centers was zero (this was obtained by adding 0.080 to -0.080). For math activities, Head Start parents in low family service worker outreach centers did .244 standard deviation units more math activities than parents in the control condition, while parents in high family service worker outreach centers did .155 standard deviation units more math activities than parents in the control condition. Thus, for both outcomes, high family service worker outreach centers were not less effective compared to not attending Head Start, but were less effective compared to low-outreach Head Start centers.
Family service worker outreach did not predict the treatment effect for the other examined parenting outcomes.

A significant and positive interaction between Head Start random assignment and requiring frequent home visiting was found for parents’ likelihood of reading daily, indicating that centers that required four or more home visits were more effective at promoting parents’ daily reading with children than centers that required three or fewer home visits (Figure 6). Specifically, parents in low outreach centers were 2.3% more likely to read to children daily compared to parents in the control group (an effect that was not significantly different from zero), while parents in high outreach centers were 10.2% more likely to read to children daily compared to the control group. Providing four or more home visits did not predict the treatment effect for any other examined outcomes.

It should be noted that the two significant interactions for parents’ likelihood of reading daily emerged even though treatment effects for daily reading were not found to vary significantly overall in Chapter 2. These findings are in line with prior work showing that even in the absence of significant treatment effect variation overall it is sometimes still possible to predict treatment impact variation with theoretically-informed site-level characteristics (Bloom & Spybrook, 2017; Sabol et al., 2019).

**Classroom Quality Results**

Neither ECERS-R nor Arnett CIS quality scores predicted the treatment effect for the examined outcomes (Table 5). The one exception is that Arnett CIS scores moderated the treatment effect for the likelihood of parents reading to children daily (Figure 7). Centers with higher Arnett scores were 7% more effective at improving the likelihood of daily reading than centers with lower Arnett scores.
Supplementary findings from analyses using four ECERS-R subscales (interactions, program structure, language and reasoning, and activities) are shown in Supplementary Table 1. None of the ECERS-R subscales significantly moderated the effect of Head Start random assignment on any of the five examined parenting outcomes. Supplementary findings from analyses using a more stringent cutoff (scores of 6 and above instead of 5 and above) for “high” ECERS-R quality are shown in Supplementary Table 2. When measured with this more stringent cut-off, ECERS-R scores still did not moderate the Head Start treatment effect on parenting outcomes.

**Supplementary Descriptive Results**

The finding that centers with high family service worker parent outreach were less effective at promoting parents’ likelihood of daily reading and their math activities than centers with lower outreach was unexpected and puzzling. I postulated that it was unlikely that family service worker outreach itself was the true cause of smaller treatment effects in high compared to lower outreach centers given strong theoretical reasons to believe otherwise. I instead reasoned that one possible alternative explanation could be that family service worker outreach was a proxy for some other characteristic of centers that made parents at those centers more or less responsive to Head Start. For example, if low parental education were associated with weaker benefits of Head Start in terms of parents’ math activities with children, and centers with high family service worker outreach tended to serve a higher percentage of families with low parental education (either randomly, or because those center directors perceived these families to be more in need of outreach), this could explain this unexpected and unintuitive finding. To explore this possibility, I examined the family demographic characteristics used as covariates for centers high versus low in family service worker outreach (see Supplementary Table 3). I found that parents
in centers with low family service worker outreach were more likely to be Black, less likely to be White, more likely to have Spanish as their home language, and were slightly older. To the extent that racial and ethnic minority parents tend to experience more daily stressors and economic strain than White parents, even among low-income families, these patterns suggest that families in centers with low family service worker outreach may have been slightly more disadvantaged than families in high outreach family service worker outreach centers. I explore the possibility that theoretically-relevant characteristics of families within centers account for treatment effect variation more thoroughly in Study 3.

**Discussion**

Understanding which Head Start center characteristics make some programs more effective than others at improving parenting outcomes is imperative for Head Start decision makers’ ability to strengthen those characteristics across programs and therefore maximize Head Start’s impact on parenting practices. Indeed, this aim is in line with Head Start’s new PFCE framework, the goal of which is to assist programs in increasing engagement with parents in order to more fully involve them in their children’s learning and development. This study attempted to uncover whether two theoretically important center characteristics for improving parenting outcomes—parent outreach and center quality—could help to explain why some centers were more effective than others. I found that the effect of parent outreach on Head Start’s treatment effect was mixed. In-center parent outreach activities—measured via lead teachers’ and family service workers’ activities aimed at promoting parents’ engagement in their children’s learning—either did not moderate the treatment effect (lead teacher outreach) or moderated the treatment effect in the unexpected direction such that high outreach centers were less effective at promoting parents’ likelihood of reading daily and their engagement in math
activities than lower outreach centers (family service worker outreach). However, outreach activities that occurred outside the center—measured via a center’s required number of home visits—was positively associated with parents’ likelihood of reading daily. Finally, I found that center quality—measured via the ECERS-R and the Arnett CIS—did not moderate treatment effects on parents’ outcomes, except in one case, where centers with higher Arnett CIS scores were more effective at promoting parents’ daily reading.

The findings that high parent outreach, compared to lower parent outreach, either did not predict or predicted lower Head Start treatment effects on parenting outcomes were unexpected, as I had hypothesized that centers with higher parent outreach would be more effective at improving parenting outcomes. Results are perhaps more understandable for lead teachers’ parent outreach, where no effects were found, but less so for family service worker outreach, where effects were in the opposite direction of what was expected. Being a lead teacher in Head Start is analogous to being a pre-k teacher in another early learning context; the word “lead” serves to distinguish lead teachers—the main teachers in Head Start classrooms—from assistant teachers, who assist the lead teacher in Head Start classrooms. Thus, while all staff are encouraged to engage families as equal partners in children’s learning, the lead teacher’s role in this effort is perhaps less straightforward than that of family service workers. Furthermore, it could be the case that lead teachers are committed to engaging families, but that they do not have very much time with parents or they do not engage with them frequently enough and thus their outreach activities are not sufficient to contribute to Head Start’s impact on changes in parenting.

Family service workers, on the other hand, are the individuals at Head Start centers explicitly responsible for engaging families and connecting them with needed services; thus, it is most surprising that treatment effects were smaller for high family service worker outreach.
centers compared to centers with lower family service outreach. Before discussing these results further, it is important to first emphasize that high family service worker outreach was not associated with smaller treatment effects for Head Start parents compared to parents in the control condition, but only compared to parents in lower family service outreach Head Start centers. Parents in high family service worker outreach centers were no more or less likely to read to children daily than parents in the control condition, and, parents in high family service worker outreach centers improved in their math activities compared to parents in the control condition. Thus, higher family service worker outreach was certainly not harmful to Head Start’s efforts to improve parenting behaviors; it was just not as effective at improving parenting for centers high versus low in family service worker outreach. Nevertheless, these findings are still puzzling and thus worthy of further investigation.

One possible explanation for the smaller treatment effects in high family service worker outreach centers is that family service worker outreach was a proxy for some other characteristic of Head Start centers that made parents at those centers more or less responsive to Head Start. Supplementary descriptive work exploring this possibility provided some evidence that families in centers with low family service worker outreach may have been slightly more disadvantaged, given that they were more likely to be Black, less likely to be White, and more likely to have Spanish as their home language—characteristics typically associated with more disadvantaged economic circumstances, even among low-income families (Burchinal et al., 2011; Padilla, Cabrera, & West, 2017; Reardon & Galindo, 2009). Head Start providers do not typically decide how they will provide services in a vacuum; rather, they tailor their services to the families they serve. Thus, it could be that family service workers at centers that serve higher proportions of disadvantaged families respond by providing more outreach activities in an attempt to provide
them additional needed support. If more disadvantaged parents are less likely to change their parenting behaviors as a result of Head Start than less disadvantaged parents, then what appears to be an effect of high family service worker outreach is in reality an effect that stems from the makeup of families served at those centers. This possibility is explored more thoroughly in Chapter 4 of this dissertation.

Another possibility is that the way that the HSIS measured parent outreach did not adequately capture the mechanism by which parent outreach changes parents’ behavior. Theoretically, in order for outreach to be effective at changing parenting, the content of messages that providers give parents must be developmentally appropriate and accurate, the outreach activities by Head Start staff must be frequent enough that parents are able to internalize the messages, and the relationships between providers and parents should be characterized by trust and mutual respect. Parents would also ideally trust and have good working relationships with providers (Barnes & Nolan, 2019; Brookes, Summers, Thornburg, Ispa, & Lane, 2006; Korfmacher, Green, Spellmann, & Thornburg, 2007; Nix, Bierman, Motamedi, Heinrichs, & Gill, 2018; Sierau, Brand, & Jungmann, 2012). Taking positive disciplinary strategies as an example, providers would ideally provide parents with effective strategies for handling misbehavior without becoming physical as well as with information about the ineffectiveness of spanking as a disciplinary strategy (as opposed to either providing misinformation on spankings’ effectiveness or even promoting the use of spanking). Providers would also ideally provide these messages frequently, and providers and parents would ideally have good relationships such that providers delivered messages with respect and sensitivity and parents trusted the providers’ expertise and intentions. It is not possible for me to take any of these dimensions of provider outreach into account, as the HSIS only includes a crude measure of whether lead teachers and
family service workers did certain outreach activities within the course of the Head Start year or not. It would have been preferable to have more information about the content of their interactions with parents, the frequency of those interactions, and quality of those interactions. Future evaluations of Head Start should include measures of these components of providers’ outreach activities.

In addition to delivering in-center outreach activities, Head Start also requires centers to provide at least two home visits per year, though centers can provide more if they choose to do so. Centers that required four or more home visits in the HSIS were more effective at improving parents’ likelihood of daily reading than were centers that required three or fewer home visits. This finding supports the hypothesis that center outreach in the form of home visiting provides parents with the kinds of direct instruction and modeling that help to improve parents’ interactions with their children. Frequent home visits did not predict variation in treatment effects for any of the other four included parenting outcomes. Home visitors typically provide a range of services to the families they serve—including parenting supports but also mental health support and connection to needed services. It is thus somewhat surprising that centers that provided more home visits were only more effective at promoting parents’ daily reading, but not parents’ overall literacy or math activities, or their harsh discipline behaviors. Perhaps with the number of objectives that home visitors are expected to address, it is easiest or most feasible for Head Start home visitors to focus on promoting daily reading—a relatively straightforward and well-established goal for parents. Head Start should consider ways to support home visitors in making the encouragement of math and other literacy activities and the discouragement of harsh punishment more straightforward as well.
As with outreach, it is possible that I did not find that home visiting moderated the treatment effect on additional parenting outcomes because the measure of home visiting available in the HSIS—the frequency of required home visits—is not the most relevant feature of home visiting with regard to how Head Start impacts parenting. For instance, it could be that the nature of the home visitor’s relationship with the parent is what matters for his or her ability to change a parent’s behavior, and it was not possible for me to discern whether families were visited by the same home visitor for every home visit or whether they were satisfied with their relationship with the home visitor. Indeed, one study found that attending more home visits was only associated with more positive parent-child interactions when parents were familiar with the home visitor (Stormshak et al., 2002). Another possibility is that requiring four home visits (which represented the lower anchor of the “high” home visiting group in the present study), though more than many centers’ requirements, is still not enough to move the dial on parents’ outcomes. An RCT of the REDI-P intervention found that frequent home visits improved parents’ reading and literacy interactions with children, but the intervention delivered 10 home visits during Head Start and another 6 after children were in kindergarten (Bierman et al., 2015). Thus, it could be that centers that require significantly more than four home visits are more effective than those that offer fewer. It was not possible for me to discern the exact number of home visits required by centers in the HSIS because only a categorical version of this variable that represented one, two to three, or three or more home visits was provided. Therefore it was not possible to explore whether providing this many home visits predicted greater Head Start treatment effects on parenting behaviors.

Centers’ classroom quality—measured via the ECERS-R and the Arnett CIS—largely did not predict variation in Head Start treatment effects. The one exception to this was that centers
with higher Arnett CIS scores were more effective at promoting parents’ daily reading than centers with lower Arnett CIS scores. The Arnett CIS measures caregiver-child interactions in the domains of teacher sensitivity, responsiveness, encouragement of children’s independence, and lower levels of punitiveness and detachment. Teachers who are more skilled in these domains may be more effective at providing direct instruction and modeling than teachers who are less skilled in these domains—it is only possible to model positive adult-child interactions if you yourself interact with children positively, and these teachers may have more skills with which to interact with parents as well.

Overall, classroom quality only predicted one outcome for one of the two available quality measures. It could be true that classroom quality is not very important for predicting Head Start’s effects on parenting outcomes. Small differences in classroom quality may simply not be apparent enough to parents to yield meaningful differences in parenting behavior at home. Also, classroom quality is largely a function of characteristics of Head Start lead teachers, and as has been discussed for the measures of outreach (including home visiting) in the HSIS, it is possible that lead teachers’ skills and engagement of parents is less important than the engagement of other program staff who might interact with parents more. It is also possible—and arguably more likely—that the measures of quality included in the HSIS were not ideal in capturing the mechanism by which quality theoretically impacts Head Start treatment effects. The ECERS-R in particular has been widely criticized for its lack of domain specificity and its poor psychometric properties (Gordon et al., 2015; Gordon, Fujimoto, Kaestner, Korenman, & Abner, 2013). The ECERS-R is a global measure of classroom quality and includes components of teachers’ relationships with students but also characteristics having to do with the resources of the centers, including available materials and the adequacy of space and furnishings. It is
possible that this scale encompassed too many disparate constructs to adequately capture any specific important characteristics having to do with Head Start quality that matter for improving parenting. Likewise, the Arnett CIS has been criticized for mostly capturing behaviors that the majority of caregivers display and thus not effectively distinguishing between high- and low-quality care (Colwell, Gordon, Fujimoto, Kaestner, & Korenman, 2013). Thus, while it is possible that classroom quality does not matter for a Head Start center’s ability to promote positive parenting, it is also likely that whether or not this is true, the measures included in the present study limited my ability to fully explore the role of classroom quality in predicting Head Start treatment impacts on parenting.

It is noteworthy that a number of center characteristics predicted differences in treatment effects for parents’ daily reading despite the finding that Head Start was consistently effective at promoting this behavior (i.e. Head Start treatment effects did not vary significantly for this outcome in Chapter 2). Finding that treatment effects do not vary significantly but that a theoretically informed variable nonetheless predicts differences in treatment impacts for that outcome is consistent with other work examining predictors of heterogeneous treatment impacts (Sabol et al., 2019). These findings do not necessarily contradict one another—even if the overall cross-site treatment effect variation is not large enough to reach statistical significance, a theoretically-informed variable could nonetheless moderate the relationship between Head Start treatment status and a given outcome such that treatment effects differ based on that variable. My findings support the notion that theoretically-informed moderators should still be explored even in the absence of significant treatment effect variation (Sabol et al., 2019).

It is important to keep in mind that the results of moderation models presented in the present study are not causal. Although families are randomly assigned to Head Start, they are not
randomly assigned to receive a given center characteristic (e.g. high frequency home visits) or not. Centers that differ in their outreach or in their quality may differ systematically in ways relevant to Head Start effects on parents. For instance, as discussed previously, it is plausible that centers that offer more family service worker outreach also serve more disadvantaged families, and disadvantage might affect a family’s likelihood of benefitting from Head Start. Likewise, centers that offer more home visits might also serve more motivated parents, and this motivation might cause parents at some centers to be more likely to increase their daily reading with children than centers serving less motivated parents. Thus, the “effect” of each center characteristic examined in the present study on centers’ ability to improve parenting outcomes should not be considered causal.

The major limitation of the present study, in addition to those mentioned in Chapter 2, is that the measures of the included center characteristics are crude and thus likely do not best capture the mechanisms by which they affect Head Start’s ability to improve parenting outcomes. As discussed previously, the measures of in-center outreach are simply counts of the number of activities that lead teachers and family service workers perform at their center. These measures do not provide any information about the frequency with which those activities occur, the quality of those provider-parent interactions, or the content of those interactions. With regard to home visiting, the measure captures the number of required home visits at the center, which may or may not reflect the actual number of home visits that families receive. Moreover, it is not known who is performing the home visits in each center. Knowing whether parents were familiar with their home visitor would have been useful, as home visiting has been shown to be most effective when parents are familiar with the home visitor (Stormshak et al., 2002). Finally, the measures of classroom quality included in the HSIS are also limited psychometrically, and may
not adequately capture specific domains of classroom quality potentially important for promoting parenting outcomes. Moreover, there is also evidence that quality scores in the HSIS are inflated, given quality ratings from other studies showing different distributions of scores, with fewer classrooms in other samples being rated as “high” quality (Mashburn et al., 2008; Perlman, Zellman, & Le, 2004).

While keeping these measurement limitations in mind, the present study makes an important contribution to the literature by being the first to explore whether theoretically important center characteristics help account for why some centers are more effective at improving parenting outcomes than others. I found that Head Start parent outreach and quality predicted some but not all parenting outcomes, and sometimes in unexpected directions. The next chapter of this dissertation extends this work by examining whether variation in the characteristics of families that individual Head Start centers serve might help explain why some centers are more effective at improving parenting outcomes than others.
Table 3

*Descriptive statistics of center characteristics*

<table>
<thead>
<tr>
<th></th>
<th>%</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>High parent outreach</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lead Teacher</td>
<td>0.227</td>
<td>0.420</td>
<td>0</td>
<td>1</td>
<td>277</td>
</tr>
<tr>
<td>Family Service Worker</td>
<td>0.498</td>
<td>0.501</td>
<td>0</td>
<td>1</td>
<td>281</td>
</tr>
<tr>
<td>Center offers 4+ home visits</td>
<td>0.210</td>
<td>0.408</td>
<td>0</td>
<td>1</td>
<td>281</td>
</tr>
<tr>
<td>High classroom quality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECERS-R</td>
<td>0.650</td>
<td>0.478</td>
<td>0</td>
<td>1</td>
<td>286</td>
</tr>
<tr>
<td>Arnett</td>
<td>0.514</td>
<td>0.501</td>
<td>0</td>
<td>1</td>
<td>286</td>
</tr>
</tbody>
</table>

*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. Ns vary across variables here because center characteristic data were not imputed and are thus missing for some centers.
Table 4
Pairwise correlations between centers’ parent outreach and classroom quality measures

<table>
<thead>
<tr>
<th>Measure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. High Lead Teacher Outreach</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. High Family Service Worker</td>
<td></td>
<td>0.209</td>
<td>***</td>
<td>--</td>
</tr>
<tr>
<td>Outreach</td>
<td></td>
<td>0.131</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. 4+ Home Visits</td>
<td>-0.031</td>
<td></td>
<td>*</td>
<td>--</td>
</tr>
<tr>
<td>4. High ECERS-R</td>
<td>-0.105</td>
<td>+</td>
<td>-0.077</td>
<td>-0.061</td>
</tr>
<tr>
<td>5. High Arnett CIS</td>
<td>0.021</td>
<td>0.001</td>
<td>0.004</td>
<td>0.443</td>
</tr>
</tbody>
</table>

***p<.001, **p<.01, *p<.05, p<.10
Table 5
*Models showing the non-causal ITT effect of Head Start parent outreach, home visiting, and classroom quality on the average Head Start effect*

<table>
<thead>
<tr>
<th></th>
<th>Child is read to daily (y/n)</th>
<th>Literacy Activities</th>
<th>Math activities</th>
<th>Child spanked in past week (y/n)</th>
<th>Number of spanks</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parent Outreach</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outreach - Lead Teacher</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>0.037</td>
<td>0.043</td>
<td>*</td>
<td>0.147</td>
<td>0.000</td>
</tr>
<tr>
<td>Treatment*High Outreach</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outreach</td>
<td>0.016</td>
<td>0.689</td>
<td>-0.011</td>
<td>0.898</td>
<td>-0.132</td>
</tr>
<tr>
<td><strong>Outreach - Family Service Worker</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>0.080</td>
<td>0.000</td>
<td>***</td>
<td>0.184</td>
<td>0.000</td>
</tr>
<tr>
<td>Treatment*High Outreach</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outreach</td>
<td>-0.080</td>
<td>0.011</td>
<td>*</td>
<td>-0.076</td>
<td>0.271</td>
</tr>
<tr>
<td><strong>Home visiting</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>0.023</td>
<td>0.191</td>
<td></td>
<td>0.152</td>
<td>0.000</td>
</tr>
<tr>
<td>Treatment*4+ Home Visits</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visits</td>
<td>0.079</td>
<td>0.047</td>
<td>*</td>
<td>-0.016</td>
<td>0.847</td>
</tr>
<tr>
<td><strong>Classroom Quality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECERS-R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>0.042</td>
<td>0.113</td>
<td>0.122</td>
<td>0.034</td>
<td>*</td>
</tr>
<tr>
<td>Treatment*High Quality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality</td>
<td>0.000</td>
<td>0.999</td>
<td>0.061</td>
<td>0.391</td>
<td>-0.030</td>
</tr>
<tr>
<td><strong>Arnett CIS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>0.007</td>
<td>0.766</td>
<td>0.157</td>
<td>0.001</td>
<td>**</td>
</tr>
<tr>
<td>Treatment*High Quality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality</td>
<td>0.070</td>
<td>0.025</td>
<td>*</td>
<td>0.008</td>
<td>0.911</td>
</tr>
</tbody>
</table>
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 effect. Covariates include those used in the original HSIS 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, number of spanks) 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.

***p<.001, **p<.01, *p<.05, +p<.10
Figure 3. Variation across centers in outreach activities.
Figure 4. Variation across centers in quality scores.
Figure 5. ITT effects of random assignment on parenting outcomes for continuous outcomes (left) and binary outcomes (right), by Family Service Worker (FSW) Outreach.

***p<.001, **p<.01, *p<.05, +p<.10
Figure 6. ITT effects of random assignment on parenting outcomes for continuous outcomes (left) and binary outcomes (right), by number of home visits.

***p<.001, **p<.01, *p<.05, +p<.10
Figure 7. ITT effects of random assignment on parenting outcomes for continuous outcomes (left) and binary outcomes (right), by Arnett CIS scores. 

***p<.001, **p<.01, *p<.05, +p<.10
CHAPTER IV: FAMILY CHARACTERISTICS AS SOURCES OF VARIATION IN HEAD START EFFECTS ON PARENTING OUTCOMES

Returning to the conceptual framework for studying sources of variation in program effects, Weiss and colleagues (2014) offer three factors that could theoretically moderate a program’s treatment effect on a given outcome: characteristics of the organization, characteristics of those participating in the intervention, and the RCT treatment contrast. The analyses presented in Chapters 2 and 3 demonstrate that there is significant variation in treatment effects across Head Start centers for multiple parenting outcomes, and that in some cases variation in treatment effects can be attributed in part to two center characteristics, parent outreach and center quality. Building on these findings, the focus of the present study is examining the second of Weiss and colleagues’ proposed sources of variation in Head Start effects on parenting outcomes: characteristics of families served by Head Start.

Different types of parents may benefit differentially from Head Start (Puma et al., 2010), and centers vary in the composition of families they serve (Tarullo et al., 2017). Centers with higher concentrations of parents who stand to benefit more from the program should be more effective than centers with lower concentrations of those types of parents. The present study examines four family-level characteristics that might make parents differentially responsive or receptive to Head Start as an intervention: household economic risk level, maternal immigrant status, family multigenerational Head Start status (whether the mother attended Head Start as a child), and parents’ initial levels of cognitive stimulation and spanking at the beginning of the Head Start year. Results from this study are important for understanding the family characteristics that contribute to Head Start treatment effects on parenting outcomes and thus can
help Head Start decision-makers understand where to target services in order to maximize Head Start’s effectiveness.

**Variation in Head Start Treatment Effectiveness by Family Characteristics**

In the intervention literature, there is long-standing acknowledgement that programs may benefit different groups of individuals differently. Program effects may be large for some groups and small or null for others even if there is a strong average effect of the program (Weiss et al., 2017, 2014). The most common characteristic to consider when hypothesizing characteristics of participants that might make them differentially able to benefit from an intervention is risk level, which is defined differently depending on the intervention and outcome in question. It should be noted that by nature of being a program for low-income families, all Head Start participants may be considered “at risk,” but even among low-income populations, there is variation in the degree to which Head Start parents exhibit risk for low levels of cognitive stimulation and high levels of harsh discipline.

There are three competing hypotheses regarding who might benefit most from interventions (Weiss et al., 2014). Specific to Head Start, the program might be better able to improve parenting for parents who enter the program at risk in some way for poorer parenting quality. These parents may be in the best position to take advantage of Head Start’s services and theoretically have the most “room to grow” in terms of improving their positive parenting behaviors over the course of the Head Start year. On the other hand, it could be that Head Start is better at improving parenting behaviors for parents who enter the program with the lowest level of risk, as they may be most equipped to take advantage of Head Start services. These parents may be most “service ready” and the most motivated to change their parenting once exposed to the direct instruction and modeling that Head Start provides. Last, it is possible that Head Start is
most effective at improving parenting behaviors for parents in the middle of these two extremes, as these families have the best mix of room for improvement and ability to take advantage of Head Start services (Weiss et al., 2014).

Another broad characteristic of families that might make them more or less likely to respond to Head Start services in terms of their parenting outcomes is their likelihood of responding positively to the direct instruction and modeling that Head Start provides. This likelihood may sometimes, but not always, overlap with characteristics that signal risk to positive parenting behaviors. For example, parents who attended Head Start themselves as children may be considered “at risk” because they grew up in poverty and have not yet escaped it. At the same time, these parents may be more receptive to Head Start as a program model and an organization and therefore their parenting may be more responsive to Head Start direct instruction and modeling.

There has been considerable empirical attention paid to answering the question of whether Head Start differentially benefits various groups of participants, but there are nonetheless significant limitations in body of this work. As with the broader Head Start literature, in comparison to those studies that consider child outcomes, very few of these studies focus on examining differential impacts of Head Start on parenting outcomes. Another issue with this prior research is that except for very few recent exceptions, studies using the HSIS and other samples to examine differential Head Start effects use interaction or subgroup models to examine whether program effects differ for different groups of families. While certainly important and informative, these studies ask whether Head Start benefits one group over another without considering the fact that these groups are not evenly distributed across centers; thus, these studies cannot differentiate variation attributable to individual families’ characteristics from variation
attributable to centers that serve different types of families. Subgroup models comparing average Head Start effects for immigrants versus non-immigrants pool all immigrants (and non-immigrants) together, ignoring the fact that they are not evenly distributed across centers. This is an important distinction because Head Start policy must be administered at the center level—Head Start cannot target individual immigrant families. Put another way, studies that employ subgroup models ask the question: Does Head Start benefit immigrants more than non-immigrants on average? Versus an approach that considers how children are served by Head Start in reality and asks: Are centers with a higher proportion of immigrants more effective compared to centers with a lower proportion of immigrants? This distinction is subtle, but important given that Head Start is delivered at the center-level.

There are a number of family characteristics that signal potential risk to poor parenting outcomes and/or receptivity to Head Start services. I chose four such characteristics to examine in the present study due to their theoretical importance for predicting variation in Head Start parenting effects or due to prior reports showing that they matter for explaining variation in Head Start effects on children and thus are of interest to explore when predicting parenting outcomes.

**Household Economic Risk**

As mentioned previously, the most commonly considered characteristic when thinking about how programs may differentially benefit individuals is “risk”. Risk is defined differently depending on the intervention or outcome of interest. Perhaps the most commonly studied risk factor is economic risk, which can be measured in a variety of ways, but ultimately signals a family’s lack of financial resources and thus difficulty meeting basic needs. Economic risk is related to poorer child academic, socioemotional, and health outcomes (Case, Lubotsky, & Paxson, 2002; Duncan & Magnuson, 2011; Evans & English, 2002; Klebanov, Brooks-Gunn,
McCarton, & McCormick, 1998; Noble, McCandliss, & Farah, 2007; Tucker-Drob, Rhemtulla, Harden, Turkheimer, & Fask, 2011; Waldfogel & Washbrook, 2011) as well as poorer quality parenting (Bradley & Corwyn, 2002; Bradley, Corwyn, Pipes McAdoo, & García Coll, 2001; Gurian, Hurst, & Kearney, 2008; Kalil, Ryan, & Corey, 2012; Kalil, Ziol-Guest, Ryan, & Markowitz, 2016). For the reasons outlined above, Head Start may be most effective at improving parenting outcomes for parents with the greatest degree of economic risk, in which case centers serving a higher proportion of at-risk families should be more effective than those serving lower proportions of high-risk families. But it could also be the case that Head Start is most effective at improving parenting outcomes for parents with low economic risk, in which case centers with the highest proportions of families with low levels of economic risk should be most effective.

There is some prior work to suggest that Head Start may benefit parents with greater economic risk more than those with lower economic risk. McCartney and colleagues (2007) found that higher quality child care (not specific to Head Start) was associated with improved maternal responsiveness, cognitive stimulation, and increased availability of learning materials in the home, and that this association was strongest for the poorest families. Specific to Head Start, a number of non-experimental studies have also found that Head Start has a greater impact on child outcomes for children experiencing greater economic risk (Cooper & Lanza, 2014; Hubbs-Tait et al., 2002; Lee, 2008, 2011; Raikes et al., 2006).

The original congressional mandate directing the department of Health and Human Services (HHS) to determine whether Head Start was effective specified that the HSIS should examine “possible sources of variation in the impact of the Head Start program.” The original analysis of the HSIS therefore included findings from subgroup analyses in which Head Start
program effects on child and parent outcomes were compared for groups differing across a number of dimensions. One of these dimensions was a composite of five characteristics signaling some degree of economic or resource-related risk: receipt of Temporary Assistance for Needy Families (TANF) or food stamps, if neither parent in the household had a high school diploma or GED, if neither parent in the household was employed or in school, if the child’s biological mother/caregiver was a single parent, and if the child’s biological mother was age 19 or younger when the child was born (Puma et al., 2010). Individually, these characteristics all signal risk to financial stability and thus risk to poorer child outcomes and to poorer parenting quality (Heinrich, 2014; Kalil, 2014; McLanahan, 2004; Morris, Gennetian, & Duncan, 2005; Ribar, 2015; Ryan, 2012; Teitler, Reichman, & Nepomnyaschy, 2004; Yeung, Linver, & Brooks-Gunn, 2002). That risk is compounded when multiple risk factors are experienced simultaneously (Ayoub et al., 2009; Burchinal, Vernon-Feagans, Cox, & Key Family Life Project Investigators, 2008; Deater-Deckard, Dodge, Bates, & Pettit, 1998; Trentacosta et al., 2008).

In the original HSIS evaluation, researchers split this household risk index into 3 categories: low/no household risk (0-2 risk factors), moderate household risk (3 risk factors), and high household risk (4-5 risk factors) and compared Head Start effects across subgroup models. The evaluation found that children in the 3-year-old cohort from families with the highest risk benefited the most with regard to multiple cognitive outcomes, but 4-year-olds did not differentially benefit depending on family risk. With regard to parenting outcomes, there was only some evidence of variation by subgroup. Specifically, in the 3-year-old cohort there were larger impacts of Head Start on reducing spanking for parents in the moderate risk group compared to both the low/no risk group and the high-risk group, but there were no differential
impacts of Head Start on parents’ reading behavior for the 3-year-old cohort, nor were there any differences for families differing on this household risk index for the 4-year-old cohort

**Immigrant Status**

Parental immigrant status is another family characteristic theoretically relevant to the effectiveness of Head Start on parenting outcomes. Immigrant parents tend to have more economic risk factors than native-born parents—on average they have fewer financial and educational resources (Pew Research Center, 2015). They are also less likely to be fluent in English (Gambino et al., 2014), which could hinder communication with Head Start staff and thus weaken the effect of Head Start on parenting outcomes compared to the effect for native-born parents, depending on how many staff at the center speak the parent’s native language (Barrueco, Smith, & Stephens, 2016; Park & McHugh, 2014; Turney & Kao, 2009). Indeed, many immigrant parents report feeling uncomfortable interacting with ECE teachers due to a lack of familiarity with U.S. educational institutions (Barrueco et al., 2016; Whatley & Batalova, 2013). Immigrant parents, particularly those who are low-income, have also been shown to engage in their children’s ECE settings less frequently (Gelatt, Adams, & Huerta, 2014; Matthews & Ewen, 2010). On the other hand, research on the “immigrant paradox” has shown that in some contexts children with immigrant parents actually perform better on health and educational outcomes than children of native-born parents (Hirschman, 2001; Palacios, 2012; Ryabov & Van Hook, 2007). Relatedly, immigrants often cite wanting better economic futures for their children than they themselves experienced in their home countries as a reason for immigrating in the first place (Perreira, Chapman, & Stein, 2006). Immigrant parents thus may be highly motivated to take advantage of Head Start services and may therefore benefit more from Head Start than other parents.
Prior non-experimental studies using subgroup models have found that Head Start effects on child outcomes are sometimes larger for children with immigrant parents, particularly for those who are Hispanic, compared to children of native-born parents (Magnuson, Lahaie, & Waldfogel, 2006). Furthermore, in their study quantifying the variation in Head Start effects on child outcomes, Bloom and Weiland (2015) found that centers with higher proportions of Dual Language Learners (DLLs) with low initial receptive vocabulary skills had larger effects on child outcomes than centers with lower proportions of low-pretest DLLs. Authors postulated that Head Start may have compensated for these children’s limited prior exposure to English. This explanation is plausible for Head Start’s effect on child outcomes, but the reasons to expect variation in effects on child outcomes may not overlap completely with the reasons to expect variation in effects on parenting outcomes. A more salient characteristic for parenting outcomes than a child’s DLL status is likely parents’ own language- and culturally-related characteristics, including being an immigrant. Immigrant parents may have more difficulty connecting with Head Start providers and services due to their often limited prior English exposure or their lack of experience with the U.S. educational system. To my knowledge, no prior study has examined whether Head Start has a greater impact on parenting outcomes for immigrant versus nonimmigrant parents.

**Multigenerational Head Start Status**

Given that Head Start has been operating for decades, it is now possible for children currently in Head Start to have parents who went to Head Start themselves as children. About 18% of former Head Start participant mothers have sent their own children to Head Start (US DOL, 2013). Participants are considered multigenerational Head Start (MGHS) families when two consecutive generations participate in the program (Chor, 2018).
Parents from MGHS families signal a unique profile of both risk for poor outcomes and receptivity to Head Start’s services. On the one hand, mothers who went to Head Start themselves as children and then send their own children to Head Start have not escaped poverty, despite their positive early educational experiences. It could be that the disadvantage associated with poverty may compound across generations such that children from MGHS families experience greater cumulative risk compared to children whose parents have experienced poverty more recently (Chor, 2018). MGHS eligibility might thus be an even stronger form of disadvantage than single-generation eligibility and thus might be associated with even poorer parenting outcomes. However, although it is true that a mother who participated in Head Start as a child was raised in poverty, many other mothers who send their children to Head Start also likely experienced disadvantage as children given the strong intergenerational cycle of poverty (Aaronson & Mazumder, 2008; Black & Devereaux, 2011; Isaacs et al., 2008; Lindahl et al., 2015). Thus, MGHS status does not on its own necessarily signal greater disadvantage compared to families in which parents did not attend Head Start as children. Moreover, compared to poor non-Head Start former participants, former Head Start participants may have experienced long-term positive outcomes, as research has demonstrated long-term benefits of Head Start on human capital outcomes such as increased high school graduation and college attendance (Carneiro & Ginja, 2014; Deming, 2009; Ludwig & Miller, 2007). Indeed, in a study examining the multigenerational effects of Head Start, compared to children of parents with lower access to Head Start as children, children of parents who had greater access to Head Start as children experienced greater long-term gains in terms of higher educational attainment and lower rates of teen pregnancy and interaction with the criminal justice system (Barr & Gibbs, 2017). Furthermore, when a child whose mother attended Head Start also attends the program, the
mother has a second chance to benefit from the Head Start services (Chor, 2018). With regard to parenting outcomes, mothers who participated in Head Start as children might be more familiar with and receptive to Head Start’s direct instruction and modeling and to the other services offered to parents. In this way, Head Start could benefit these mothers more than those that did not attend Head Start as children.

This possibility has not been previously examined, but one study has examined differences in Head Start effects on child outcomes for single versus MGHS families using the HSIS. It found that Head Start had larger and more positive impacts on cognitive and socioemotional development through third grade among the children from MGHS families compared to children from single-generation Head Start families (Chor, 2018). This study also found that at baseline, MGHS families engaged in more cognitive stimulation (including both literacy and mathematics-related activities) with children than other families. This difference in baseline cognitive stimulation (as well as differences in family resources) between single and MGHS families helped to explain differences in program impacts for children’s outcomes.

These results show that MGHS families had higher baseline cognitive stimulation, which helped explain why Head Start was more effective for their children. But this study did not actually examine whether MGHS families’ parenting practices improved more for MGHS families compared to single-generation families as a result of their children attending Head Start, though it is possible that MGHS parents do benefit more from their children attending Head Start for the reasons outlined above. In this case, centers serving a higher percentage of MGHS families should be more effective at promoting positive parenting behaviors than centers serving fewer such families.
Initial Parenting Behaviors

One last characteristic to consider that could make parents differentially likely to benefit from Head Start is their own parenting starting point. Parents who at the beginning of Head Start engage less frequently in cognitively stimulating activities with children theoretically have more potential to benefit from Head Start exposure than parents who are already engaging frequently in stimulating activities with children. For example, a mother who does not read to her child every night or who does not challenge her child linguistically is more likely to increase those cognitively stimulating behaviors after seeing a child care provider doing them than is a mother who already engages in this stimulation regularly. Similarly, a mother who is struggling with finding effective means of disciplining a child and frequently resorts to hitting or spanking is likely to be more affected by witnessing a provider stop her child’s misbehavior without resorting to physical means than a mother who never or rarely physically disciplines her child. At the same time, however, it could also be that parents who begin the Head Start year with the least supportive parenting are the most resistant to change (Ansari et al., 2016). Parents who engage in harsh punishment such as spanking often do so because they firmly believe that harsh discipline is effective and necessary, and, as a result, they may be less open to changing these behaviors (Ansari et al., 2016; Vittrup, Holden, & Buck, 2006). By contrast, parents who begin the Head Start year with the most cognitively stimulating parenting have already demonstrated high motivation to be involved in their children’s learning, and so may be more receptive to Head Start’s message and efforts to provide direct instruction and modeling. These parents may have less “room to grow”, but may nonetheless still grow more compared to other parents.

The existing literature surrounding this question is sparse, but one study suggests that this phenomenon might depend on the parenting domain under consideration. The one study to
examine this possibility used the HSIS and found that compared to children in the control condition, parents whose children were randomly assigned to Head Start improved their frequency of reading to and levels of cognitive stimulation with children regardless of their initial levels of those behaviors (Ansari et al., 2016). For discipline behavior, however, the same study found that changes in spanking as a result of Head Start depended on initial spanking levels—parents who spanked twice a week or more, but not parents that never spanked or who spanked once a week, demonstrated reductions in spanking frequency in the following year (Ansari et al., 2016). These results demonstrate different treatment effects at the individual level for parents differing in their initial parenting behavior, at least for discipline. It is thus possible that centers with different average initial parenting behaviors are differentially effective at promoting positive parenting behaviors.

The Present Study

The present study explores whether family characteristics help to explain differences in Head Start treatment effects on parenting outcomes. Specifically, I ask whether centers are differentially effective based on differences in their compositions of families at moderate-high household economic risk, who are immigrant parents, who are MGHS families, and who have different initial parenting practices.

Given that there is very little research with which to guide my expectations and given that each of these risk factors could make parents either more or less likely to benefit from Head Start, I do not offer any specific hypotheses. Rather, the questions posed in this study are more exploratory in nature.

Methods
Data and Sample

Data again come from the HSIS. Sample restrictions were made in the same way as they were for Studies 1 and 2, with the exception that samples were further restricted to cases with non-missing data on the family characteristic under investigation. This process resulted in sample sizes that ranged from 3,342-3,520 children within 281-296 Head Start Centers.

Measures

All parenting measures and covariates in the present study were identical to those included in Chapters 2 and 3. The following family characteristics were also included in this study. Note that these family characteristics are examined at the center-level, not the parent-level. That is, I examine the percentage of parents with these different characteristics (or the average across parents for initial cognitive stimulation) within centers as my moderators of interest. Conceptually this is important because Head Start programs respond to the needs of the entire parent community that they serve, rather than to the needs of individual families. Methodologically this is important because treatment effect variation is examined across centers rather than across individuals.

Household economic risk. The HSIS dataset includes a constructed variable signaling a family’s household economic risk, which is a sum of the following five variables: whether the household received food stamps or TANF in Fall 2002; if neither parent was a high school graduate; if neither parent was working; if the mother was a teen mother; and if the mother was a single mother. This constructed variable was dichotomized for each family to represent whether the family was considered “low risk” (they experienced 0, 1, or 2 of these risk factors) or “moderate-high risk” (they experienced 3, 4, or 5 of these risk factors). This variable represents
the percentage of families who were considered moderate- or high-risk within each Head Start center.

**Maternal immigrant status.** Parents reported on their immigrant status (see Study 1 Measures section). This variable represents the percentage of families with immigrant mothers within each Head Start center.

**Multigenerational Head Start family status.** Parents reported on each person living in the household, and for each individual responded to the item “Did this person ever (including currently) attend Head Start or Early Head Start?”, which allowed for identification of mothers who participated in Head Start as children. This variable represents the percentage of families with mothers who participated in Head Start as children within each center.

**Initial parenting behaviors.** Centers’ average initial cognitive stimulation is a center-level average of parents’ Fall cognitive stimulation scores (see Study 1 Measures section) for each center. Centers’ proportion of families initially spanked represents the percentage of families within each center who reported spanking their child in Fall 2002 (see Study 1 Measures section).

**Analytic Strategy**

To explore whether each family characteristic moderated the Head Start treatment effect on each parenting outcome (i.e. whether they predict variation in treatment effects), I estimate the same model presented in Chapter 3 except that at level 2 instead of entering in a center characteristic, I enter in each family characteristic, aggregated at the center level. Specifically, I estimate the following model:

*Level 1: Parents*

\[
Parenting_{T2ij} = \alpha_j + \beta_j RA + \lambda_i Parenting_{T1} + \sum_{k=1}^{K} \pi_k X_{ki} + \varepsilon_{ij}
\]  

[Eq. 1]
Level 2: Head Start Centers

\[ \alpha_j = \alpha_j \]  \hspace{1cm} [Eq. 2]

\[ \beta_j = \beta_0^* + \pi FamilyCharacteristic_j + r_j \]  \hspace{1cm} [Eq. 10]

where all parameters in Eq. 1 and 2 are defined in the same way as in Chapter 2; \( \beta_j \) is a slope that varies randomly across centers and represents the average effect of random assignment for Head Start center \( j \); \( \beta_0^* \) represents the grand mean Head Start ITT effect size for centers where the value of the center-level family characteristic in the model is zero; \( \pi \) is a slope for the non-causal “effect” of each center-level family characteristic on the mean Head Start treatment effect. This estimate represents the rate of change in the mean Head Start ITT effect size per point (or percentage point) increase in the center-level family characteristic in the model.

In these models, \( \pi \) is the main parameter of interest because it indicates whether and how much each family characteristic contributes to heterogeneity in the treatment effect. The statistical significance of \( \pi \) was tested using the corresponding \( z \) statistic from the 2-level mixed model, which was obtained using the MIXED command in Stata 15.0.

Results

Descriptive Results

Descriptive results for family characteristics included as moderators in the present study are shown in Table 6. Within centers, on average 25.4% of families demonstrated moderate-high economic risk, 18.7% of families had an immigrant mother, and 24% of families were MGHHS families (they had mothers who had attended Head Start as children). Within centers, parents’ initial cognitive stimulation scores ranged from 0.308 to 2, with an average score of 1.156, and on average 41.6% of families initially reported spanking their children in the past week. Variation across centers in each of these family characteristics is shown in Figure 8.
Pairwise correlation coefficients between each of the center-level family characteristics included in the present study are displayed in Table 7. Correlations between family characteristics were small to moderate. Results elucidate a pattern by which families with moderate-high economic risk, MGHS families, families who are cognitively stimulating, and families who spank children are somewhat likely to be served by the same center (as evidenced by the positive correlations between these variables), but those same centers are less likely to serve a high proportion of immigrant families. Specifically, the proportion of moderate-high risk families, the proportion of MGHS families, parents’ average initial cognitive stimulation, and parents’ initial spanking use were all positively correlated with one another ($r$s ranged from .11 to .35), while the percentage of immigrant families served was negatively correlated with each of these family characteristics ($r$s ranged from -.11 to -.34).

This pattern can also be seen in Supplementary Tables 4-8, where the value of each family-level characteristic as well as all covariates were stratified by each family-level characteristic. The top panel of each of these tables (labeled “Family Moderators”) demonstrates that when centers serve higher proportions of immigrants, they serve lower proportions of families with moderate-high economic risk, lower proportions of MGHS families, families with lower average initial cognitive stimulation scores, and lower proportions of families who initially spanked—and vice versa.

**Household Economic Risk Results**

Results from models testing whether each family characteristic moderated the effect of Head Start random assignment on each parenting outcome are presented in Table 8. The “Treatment” coefficients in the first row of Table 8 under “Family Risk Composite” represent the estimated intercept, $\hat{\beta}_0^*$ from Eq. 10, the grand mean ITT effect size for centers with zero
moderate-high risk families. The “Treatment*High Family Risk” coefficients represent \( \pi \) from Eq. 10, the rate of change in the mean Head Start ITT effect size per percentage point increase in moderate-high risk families within a center. These coefficients are similarly defined for the other center-level family characteristics included in the study.

Centers’ percentage of moderate-high risk families did not predict the treatment effect for any of the examined parenting outcomes.

**Immigrant Status Results**

Centers serving more immigrant families were less effective at reducing the likelihood of parents’ spanking than were centers serving fewer immigrant families (Figure 9). More specifically, the ITT effect size is -0.048 for parents in centers serving 0% immigrant families (which represents 171 centers), meaning that in these centers, the likelihood of spanking is reduced by 4.8%—a positive outcome. For centers serving 100% immigrants (representing 13 centers), however, the ITT effect size increases by 0.129 to 0.081 for centers serving (meaning that in these centers, the likelihood of spanking increases to 8.1%). Centers’ percentage of immigrant families did not predict the Head Start treatment effect for the other examined parenting outcomes.

**Multigenerational Head Start Status Results**

Centers serving more MGHS families were more effective at improving parents’ math activities with children over the course of the Head Start year than were centers serving fewer MGHS families (Figure 10). The grand mean ITT effect size for math increased by 0.433 standard deviation units when comparing centers serving 0% to 100% MGHS families (representing 112 and 9 centers, respectively). Centers’ percentage of MGHS families did not predict the treatment effect for the other examined parenting outcomes.
Initial Parenting Results

Centers serving parents with higher average initial cognitive stimulation scores were more effective at improving parents’ likelihood of daily reading with children and their overall literacy activities than centers serving parents with lower initial cognitive stimulation scores (Figure 11). For every point increase (on a 2-point scale) in a center’s average initial cognitive stimulation score, the treatment effect for likelihood of daily reading increased by 20.2% and the treatment effect for the frequency of literacy activities increased by 0.350 standard deviation units. Centers’ initial cognitive stimulation did not predict the Head Start treatment effect for the other examined parenting outcomes. Centers’ proportion of families who initially spanked children did not predict the Head Start treatment effect for any of the examined parenting outcomes.

Discussion

The goal of the present study was to examine whether Head Start centers differ in their effectiveness at improving parenting outcomes based on four theoretically relevant characteristics of families: their household economic risk, immigrant status, MGHS status, and initial parenting practices. If certain types of families are more or less likely to benefit from Head Start, then differences in the composition of families at sites should contribute to differences in effectiveness across centers. I found that neither a composite measure of household economic risk nor parents’ initial spanking use predicted variation in Head Start effects on parenting outcomes, but that immigrant status, MGHS status, and parents’ initial cognitive stimulation did. Specifically, centers serving a higher percentage of immigrant families were less effective at reducing parents’ use of spanking; centers serving a higher percentage of MGHS families were more effective at improving parents’ math activities with children, and centers serving families
with higher average initial cognitive stimulation scores were more effective at promoting parents’ daily reading and overall literacy activities with children.

It was surprising that household economic risk did not predict variation in Head Start treatment impacts on any of the examined parenting outcomes. McCartney and colleagues (2007) found that it was the poorest parents who benefited most from high-quality ECE in the realms of maternal responsiveness, cognitive stimulation, and availability of learning materials in the home. However, by virtue of being eligible for Head Start, all families in the present study were poor. Perhaps the additional components of economic risk measured by the HSIS composite measure of household economic risk (TANF or food stamps receipt, neither parent had a high school degree or GED, neither parent was employed or in school, the child’s mother was a single parent, and the child’s mother was a teenager) did not provide any additional ability to predict parents’ responsivity to Head Start, and by virtue of being low-income, Head Start parents all share a sufficiently similar degree of economic risk. In the HSIS final report, parents with children in the 3-year-old cohort who were in the “moderate” risk group (those who experienced three of the five risk factors included in the composite) reduced their spanking compared to parents in the control group more than did parents in the other two risk groups. Given that I did not distinguish moderate from high risk families and that I combined cohorts, perhaps it is less surprising that my findings differ from those of that report. Regardless, my results do not provide evidence that household economic risk – as measured by this particular composite measure in the HSIS – predicted Head Start’s effects on parenting outcomes in either direction. Other family characteristics in the present study, however, did.

The percentage of immigrants served by a center did not predict variation in Head Start treatment effects for parents’ cognitively stimulating activities with children, but centers serving
higher percentages of immigrant families were less effective at reducing parents’ use of spanking compared to centers serving lower percentages of immigrant families. Specifically, centers serving no immigrants were somewhat effective at reducing parents’ use of spanking, but the more immigrants a center served, the less effective centers became at reducing spanking use, and centers serving all immigrants actually increased their use of spanking more compared to control group parents. It is notable that across the moderators examined for Chapters 3 and 4, the percent of immigrants served at a center was the only center or family characteristic that predicted variation in Head Start’s effect on parents’ spanking use. It is difficult to explain this finding given that immigrants are a heterogeneous group representing many backgrounds and cultures, and it was beyond the scope of the present investigation to examine effects by immigrants’ region of origin. However, the majority of immigrants in the sample were from Spanish-speaking countries, given that the majority of immigrants were Hispanic, spoke Spanish at home, and had children tested in Spanish. Compared to White parents, Hispanic parents have been shown to be more likely to employ spanking as a disciplinary strategy (Gershoff, Lansford, Sexton, Davis-Kean, & Sameroff, 2012). Thus, perhaps this finding in part reflects a stronger commitment to spanking for Hispanic immigrant parents compared to other groups. It is not clear, however, why Head Start attendance would be associated with an increase in the incidence of spanking use for this group. The original HSIS final report did not find that Head Start was associated with children’s behavior problems (though they did not examine this association for immigrants specifically), so it is likely not the case that increases in spanking resulted from increases in children’s challenging behaviors. Should these findings be replicated in other samples, Head Start should target efforts to reduce spanking towards centers serving more immigrant families, and future work should attempt to uncover the mechanism by which Head Start is associated
with increases in spanking for immigrant families, as spanking increases the risk for more serious abuse (Brown et al., 1998; Sedlak et al., 2010) and thus should be prevented.

It is somewhat easier to postulate reasons for the null effects of Head Start random assignment on parents’ cognitively stimulating activities. Hispanic parents—who made up most of the immigrants in the HSIS—have traditionally viewed education as the school’s, rather than a parent’s, responsibility (Barrueco et al., 2016) and thus may not have been responsive to Head Start’s messages regarding their role in supporting their children’s learning. Immigrant parents are also typically less familiar with the U.S. education system than native-born parents (Barrueco et al., 2016; Whatley & Batalova, 2013) and have lower levels of fluency in English (Gambino et al., 2014). Immigrant parents, particularly those with limited English language skills, have also reported feeling unwelcome in educational settings (Adams & McDaniel, 2009; Park & McHugh, 2014; Turney & Kao, 2009). All of these factors may contribute to immigrant parents engaging less in their children’s Head Start centers compared to their native-born peers (Barrueco et al., 2016), and thus make them less likely to benefit from the direct instruction and modeling that those centers may provide. Centers may need to use different strategies than they do for other families to effectively engage immigrant parents in their children’s learning. For example, programs might consider improving providers’ understanding of cultural differences affecting parent engagement through professional development (DeJaeghere & Cao, 2009; Trumbull, Rothstein-Fisch, & Hernandez, 2003) or hiring more bilingual program staff to engage immigrant families (Moore, Caal, Lawner, Rojas, & Walker, 2014; Starkey & Klein, 2000; Tang, Dearing, & Weiss, 2012), as these strategies are associated with greater feelings of being welcome in ECE settings and of parent engagement among immigrants.
Whereas serving more immigrant families predicted only parents’ disciplinary behaviors with children, the percentage of MGHS families served by a center and parents’ average initial cognitive stimulation in a center predicted Head Start’s effect on parents’ cognitively stimulating behaviors with children. Specifically, centers serving a higher percentage of MGHS families were more effective at improving parents’ frequency of math activities with children and centers serving families with higher average initial cognitive stimulation scores were more effective at improving parents’ likelihood of daily reading and their frequency of literacy activities with children. MGHS families experience a unique combination of both potential risk to poor parenting outcomes but also potential receptivity to Head Start’s efforts to improve parenting. These parents were raised in poverty and have not escaped it – signaling risk – but also may have a deeper trust in Head Start as an institution and may be more receptive to its goals given that they have now been doubly exposed to it. That centers with higher percentages of MGHS families were better at improving parents’ math activities with children is a noteworthy finding given the importance of early math learning for future achievement (Claessens, Duncan, & Engel, 2009; Duncan et al., 2007; Heckman, 2008; Ritchie & Bates, 2013; Romano, Babchishin, Pagani, & Kohen, 2010). Compared to reading activities, parents tend to do fewer math activities with children given their own discomfort with math (Levine, Suriyakham, Rowe, Huttenlocher, & Gunderson, 2010). Compared to other low-income parents, though, MGHS parents may have had greater initial comfort with math compared to other low-income parents, given Head Start’s long-term positive outcomes on educational attainment (Barr & Gibbs, 2017). This possible greater initial comfort may have translated into greater improvements in parents’ math activities with children and thus greater Head Start effects in this domain, much the same way that centers
with higher average initial cognitive stimulation scores were more effective at promoting parents’ reading and overall literacy activities with children (see below).

The finding that centers who served parents with higher average initial cognitive stimulation scores were more effective at promoting parents’ reading and other literacy activities with children supports the hypothesis that parents who are most “service ready” with regard to parenting behaviors are also those most likely to improve their parenting as a result of Head Start. These parents may have been more highly motivated to be involved in their children’s learning than parents who less frequently did cognitively stimulating activities with children and thus may have been most receptive to further time investments in these behaviors. The cognitive stimulation measure at baseline was made up mostly of literacy as opposed to math activities; thus these parents already had some proclivity towards literacy-related activities that they were able to further improve over the Head Start year. If there had been a better measure of parents’ initial engagement in math activities available, perhaps I would have also found that centers with higher levels of initial math activities were more effective at promoting parents’ math activities with children. The one prior study to examine parents’ gains from Head Start as a function of their initial parenting skills found that parents whose children were randomly assigned to Head Start improved their frequency of reading to and levels of cognitive stimulation with children regardless of their initial levels of those behaviors (Ansari et al., 2016). However, this study examined this question at the individual family level and did not ask whether centers were more or less effective based on their families’ average initial parenting behaviors. Perhaps there is something about being around other highly engaging and enriching parents during times that parents engage with their children’s Head Start centers that promotes parents’ own stimulating activities with children and thus centers’ effectiveness in this domain. The importance of social
support for parents’ engagement and wellbeing, including within ECE settings, supports this supposition (Bolívar & Chrispeels, 2011; Small, 2009).

While parents’ average initial cognitive stimulation within a center boosted centers’ effects on some parenting outcomes, the average proportions of parents within a center who initially spanked children did not. This was somewhat surprising given that in their study of child-level effects of Head Start on parenting, Ansari and colleagues (2016) found that only parents with the greatest initial levels of spanking reduced their spanking frequency after the Head Start year. Again, though, this study examined changes in parenting at the family level, not the center level. At the center level, it seems that few characteristics—apart from the percentage of immigrants served—influence Head Start’s ability to lower parents’ use of spanking. Future research should work to uncover the extent to which Head Start centers actively attempt to lower parents’ spanking use, for it could be that Head Start providers do not communicate with parents about disciplinary strategies due to the sensitive nature of the topic or the widely held belief that discipline is a parent’s decision and should not be intervened upon.

At the center-level, household economic risk status, MGHS status, initial cognitive stimulation, and initial spanking use were all positively correlated. To the extent that these characteristics moderated the Head Start treatment effect on parenting outcomes, they promoted parents’ cognitively stimulating behaviors with children. At the same time, immigrant status and parents’ initial spanking use were also modestly correlated, and the proportion of immigrant families at a center was the only characteristic across studies to predict Head Start’s effects on parents’ spanking use. While these groups of findings are certainly not definitive, they suggest that families that share certain characteristics may be clustered at centers together. Future work should perform factor analysis to identify groups of parents that tend to be clustered together in
order to explore whether a center’s composition of certain profiles of families are differentially effective at improving parenting behaviors.

The main goal of the present study was to uncover family characteristics that help predict differences in Head Start’s treatment effects on parents in order to help identify populations of families that Head Start decision makers might want to target. What my findings say about the types of centers that Head Start should target for services depends on the goal. If the goal is to target centers that serve families who stand the best chance of benefitting from Head Start, thereby improving overall Head Start effectiveness, then my results suggest that centers should target areas that serve high proportions of MGHS families and families who are already engaging frequently in cognitively stimulating activities with their children when they begin Head Start, for these centers stand to make the most gains in the realms of reading daily, overall literacy activities, and math activities. If, however, the goal is to target families who need the most support due to being less likely to benefit from Head Start, then findings suggest that Head Start should target centers serving high proportions of immigrant families, for they increased their use of spanking as a result of Head Start. Of course, findings would need to be replicated and considered within the context of the broader Head Start literature before firm recommendations could be made.
### Table 6

**Descriptive statistics of center-level family characteristics**

<table>
<thead>
<tr>
<th></th>
<th>Mean/%</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of moderate-high risk families</td>
<td>0.254</td>
<td>0.241</td>
<td>0</td>
<td>1</td>
<td>296</td>
</tr>
<tr>
<td>Percentage of families with an immigrant mother</td>
<td>0.187</td>
<td>0.276</td>
<td>0</td>
<td>1</td>
<td>296</td>
</tr>
<tr>
<td>Percentage of multigenerational Head Start families</td>
<td>0.240</td>
<td>0.264</td>
<td>0</td>
<td>1</td>
<td>283</td>
</tr>
<tr>
<td>Average initial cognitive stimulation score(^1)</td>
<td>1.156</td>
<td>0.248</td>
<td>0.308</td>
<td>2</td>
<td>295</td>
</tr>
<tr>
<td>Percentage of families who initially spanked</td>
<td>0.416</td>
<td>0.272</td>
<td>0.000</td>
<td>1</td>
<td>295</td>
</tr>
</tbody>
</table>

\(^1\)This variable is mean-centered in analyses

*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. Ns vary across variables because center characteristic data were not imputed and are thus missing for some centers.
Table 7  
*Pairwise correlations between center-level family characteristics*

<table>
<thead>
<tr>
<th>Measure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Percent Moderate-High Family Risk</td>
<td>--</td>
<td></td>
<td>***</td>
<td>--</td>
</tr>
<tr>
<td>2. Percent with Immigrant Mother</td>
<td>-0.209</td>
<td>***</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>3. Percent Multigenerational Head Start</td>
<td>0.156</td>
<td>**</td>
<td>-0.336</td>
<td>***</td>
</tr>
<tr>
<td>4. Average Initial Cognitive Stimulation</td>
<td>0.108</td>
<td>+</td>
<td>-0.299</td>
<td>***</td>
</tr>
<tr>
<td>5. Percent who Initially Spanked</td>
<td>0.157</td>
<td>**</td>
<td>-0.107</td>
<td>+</td>
</tr>
</tbody>
</table>

***p<.001, **p<.01, *p<.05, +p<.10
Table 8
Models showing the non-causal ITT effect of Head Start centers’ average economic risk status, immigrant status, multigenerational Head Start status, and initial parenting on the average Head Start effect

<table>
<thead>
<tr>
<th>Child is read to daily (y/n)</th>
<th>p</th>
<th>Literacy Activities</th>
<th>p</th>
<th>Math activities</th>
<th>p</th>
<th>Child spanked in past week (y/n)</th>
<th>p</th>
<th>Number of spanks</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Family Risk Composite</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>0.062</td>
<td>0.015 **</td>
<td>0.152</td>
<td>0.005 **</td>
<td>0.120</td>
<td>0.029 *</td>
<td>0.002</td>
<td>0.934</td>
<td>0.002</td>
</tr>
<tr>
<td>Treatment*High Family Risk</td>
<td>-0.072</td>
<td>0.336</td>
<td>0.002</td>
<td>0.991</td>
<td>0.231</td>
<td>0.150</td>
<td>-0.096</td>
<td>0.204</td>
<td>-0.013</td>
</tr>
<tr>
<td><strong>Immigrant Status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>0.054</td>
<td>0.006 **</td>
<td>0.180</td>
<td>0.000 ***</td>
<td>0.221</td>
<td>0.000 ***</td>
<td>-0.048</td>
<td>0.015 *</td>
<td>-0.004</td>
</tr>
<tr>
<td>Treatment*Immigrant</td>
<td>-0.063</td>
<td>0.314</td>
<td>-0.150</td>
<td>0.264</td>
<td>-0.210</td>
<td>0.123</td>
<td>0.129</td>
<td>0.040</td>
<td>0.016</td>
</tr>
<tr>
<td><strong>Multigenerational Head Start Status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>0.035</td>
<td>0.121</td>
<td>0.108</td>
<td>0.025 *</td>
<td>0.079</td>
<td>0.099 +</td>
<td>-0.004</td>
<td>0.857</td>
<td>0.000</td>
</tr>
<tr>
<td>Treatment*Mother in HS</td>
<td>0.019</td>
<td>0.770</td>
<td>0.172</td>
<td>0.223</td>
<td>0.433</td>
<td>0.002 **</td>
<td>-0.080</td>
<td>0.237</td>
<td>-0.005</td>
</tr>
<tr>
<td><strong>Initial Cognitive Stimulation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>0.042</td>
<td>0.007 **</td>
<td>0.158</td>
<td>0.000 ***</td>
<td>0.183</td>
<td>0.000</td>
<td>-0.021</td>
<td>0.200</td>
<td>-0.001</td>
</tr>
<tr>
<td>Treatment*Initial Stimulation</td>
<td>0.202</td>
<td>0.010 *</td>
<td>0.350</td>
<td>0.039 *</td>
<td>0.222</td>
<td>0.200</td>
<td>0.023</td>
<td>0.775</td>
<td>-0.001</td>
</tr>
<tr>
<td><strong>Initial Spanking</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>0.067</td>
<td>0.059 +</td>
<td>0.194</td>
<td>0.010 *</td>
<td>0.133</td>
<td>0.078 +</td>
<td>-0.026</td>
<td>0.471</td>
<td>-0.002</td>
</tr>
<tr>
<td>Treatment*Initial Spanking</td>
<td>-0.055 0.434</td>
<td>-0.096 0.520</td>
<td>0.101 0.503</td>
<td>0.009 0.903</td>
<td>0.002 0.911</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*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 effect. Covariates include those used in the original HSIS 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, number of spanks) 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.

***p<.001, **p<.01, *p<.05, +p<.10
Figure 8. Variation across centers in household economic risk, maternal immigrant status, multigenerational Head Start status, and parents’ initial cognitive stimulation and spanking.
Figure 9. ITT effect of random assignment on parenting outcomes for continuous outcomes (left) and binary outcomes (right), by percentage of families within a center with an immigrant mother. ***p<.001, **p<.01, *p<.05, +p<.10
Figure 10. ITT effect of random assignment on parenting outcomes for continuous outcomes (left) and binary outcomes (right), by percentage multigenerational Head Start (MGHS) families within a center.

***p<.001, **p<.01, *p<.05, +p<.10
Figure 11. ITT effect of random assignment on parenting outcomes for continuous outcomes (left) and binary outcomes (right), by pre-academic stimulation scores within a center. Effects are shown for centers with the average center-level pre-academic stimulation, and for those with pre-academic stimulation scores 1 standard deviation below and 1 standard deviation above the average.

***p<.001, **p<.01, *p<.05, *p<.10
CHAPTER V: GENERAL DISCUSSION

Parents have a profound impact on their children. The differences in parenting that children from more versus less advantaged families receive have important implications for differences in children’s skill development—differences that begin early and persist throughout the life course. The present dissertation makes an important contribution to our understanding of the potential for children’s ECE programs to impact the parenting that they receive by expanding our knowledge of the potential for Head Start to impact positive parenting behaviors. Prior work has focused on the average effect of Head Start on parenting, ignoring vast variation in how Head Start centers operate and thus the potential for variation in program effectiveness across Head Start centers. In this dissertation, I presented a series of studies in which I quantified variation in Head Start impacts across centers on parents’ cognitively stimulating and disciplinary interactions with their children and then examined theoretically relevant center and family characteristics as potential predictors of differences in effectiveness across centers. The first study established that there was significant variation in Head Start effects on parents’ literacy and math activities and to a lesser extent, parents’ use and frequency of spanking. The second study found that some center characteristics predicted variation in Head Start effects on parents’ outcomes, but sometimes in the unexpected direction. The third study found that some characteristics of the families that centers serve also predicted variation in Head Start effects on parents’ outcomes, in both positive and negative directions.

Variation in Head Start Treatment Effects on Parenting Outcomes

That Head Start centers vary significantly in the extent to which they are able to promote parents’ positive interactions with children is an important finding worthy of further consideration. Head Start 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. This system is distinct from most other publicly funded programs, where federal funding is typically distributed to individual states which then oversee funding to more local grantees, or where funding originates from the state level. In this more typical system, programs within states may be operated fairly similarly given common state requirements and low levels of freedom with regard to adhering to those requirements. By contrast, Head Start grantees are subject to federal requirements and standards, but have a high degree of freedom in terms of how they wish to go about meeting those standards. As such, individual Head Start programs vary considerably with regard to the programming they offer, the curricula they use, and the ways in which they choose to engage families. 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 this variation has meaningful consequences for 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, despite Head Start centers’ common funding source, all Head Start centers should not be considered equal with regard to effectiveness on children or on parents.

**Head Start and Parents’ Cognitive Stimulation**

Head Start effects on parenting outcomes were shown to be largest on average and also to vary most for parents’ cognitively stimulating activities with children. Head Start’s main overarching goal is to promote children’s school readiness. This dissertation finds that Head Start is effective overall at promoting parents’ contributions to this goal, yet centers nonetheless vary in their ability to do so. And, importantly, some centers are less effective than their local
alternatives at promoting parents’ cognitively stimulating behaviors, meaning that those parents may have improved more had their children not been assigned to or had they not attended Head Start. There is therefore still room for improvement in this domain. Across studies, centers that required more home visits, centers that were higher in caregiver-child interaction quality as measured by the Arnett CIS, centers that served relatively more MGHS families, and centers that served families who initially engaged in relatively high levels of cognitive stimulation were more effective at promoting parents’ cognitively stimulating interactions with children than their alternatives. To the extent that these characteristics can be grouped, these findings together suggest that centers with greater capacity for engaging families and centers serving families with relatively high motivation and receptivity to Head Start’s messaging and goals are better able to promote parenting activities known to support children’s learning than those with lower capacity and receptivity.

It is noteworthy that it was in some ways the strongest centers—with regard to both program and family characteristics—that had the greatest effects on parents’ cognitive stimulation. Programs that actively seek to support parents through frequent home visits appear to be more effective than other centers at promoting daily parent-child reading, an important routine to establish early in a child’s life. This finding supports the notion that active and frequent engagement of parents in their children’s learning is an important component for a center’s ability to promote these desired parenting behaviors. It is also noteworthy that it was programs higher in quality specific to caregiver-child interactions (as measured by the Arnett CIS)—but not centers higher in overall quality (as measured by the ECERS-R)—that were most effective at promoting parents’ daily reading. Providers at these centers were perhaps better able to communicate with parents and to model positive adult-child interactions. If replicated with
better measures of classroom quality such as the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2007), this finding suggests that promoting providers’ ability to engage with children in sensitive and responsive ways may benefit children not only directly through interactions with these caregivers, but also potentially indirectly through daily reading with parents. Efforts to support ECE teacher quality may thus be multifaceted.

With regard to characteristics of families, parents who attended Head Start themselves as children have had more exposure to the program and thus may trust and be more receptive to Head Start’s services than other parents. This greater trust and receptivity may have resulted in greater effectiveness of centers serving high proportions of MGHS families. Similarly, parents who frequently engage in cognitively stimulating activities with their children at the beginning of the Head Start year have already shown commitment to their children’s learning and are perhaps more motivated to improve upon that commitment as a result of their interaction with the program. This finding suggests that when parents’ cognitively stimulating behaviors begin earlier, parents are better able to benefit from services. Early Head Start programs should thus continue their attempts to bolster parents’ early engagement in their young children’s learning so that parents are well equipped to build upon that foundation once their children enter Head Start.

In the case of both MGHS and initial cognitive stimulation levels, larger effects for centers serving these families may be due to greater impacts for individual parents that add up to result in greater center-level effects by nature of there being a high percentage of those types of parents in a center. But it could also be that there is something about being in a center where parents are more motivated and capable that changes the center’s overall climate and makes all parents within that center more likely to benefit from Head Start. Such compositional effects are
intriguing and point to the need for future research to consider the role that center climate plays in centers’ effectiveness.

While some center and family characteristics predicted larger Head Start effects on parenting outcomes, centers in which family service workers did more parent outreach activities over the course of the Head Start year were unexpectedly less effective at promoting parents’ cognitively stimulating interactions with children. This finding should not be taken to mean that parent outreach, broadly speaking, is detrimental to a center’s ability to promote parents’ interactions with their children. First, centers with high family service worker outreach were more effective when compared to the control group, signaling that at a minimum family service worker outreach was not detrimental to Head Start’s goal of promoting parents’ outcomes. Second, family service workers respond to the needs of families at their centers and thus likely provide more outreach activities when families are in greater need of them. If families who are in greater need of Head Start services are less likely to benefit from the program (as the results with regard to MGHS families and families with greater initial cognitive stimulation suggest), then what appears to be an effect of family service worker outreach could really be an effect of greater need or lower responsivity to services. Finally, the measure of parent outreach available in the HSIS is extremely limited—it is a count of the number of activities that family service workers did over the course of the Head Start year and does not provide any information about the content, frequency, or quality of those interactions.

This unexpected finding points to the need for more research, ideally with better parent outreach measures, on the effectiveness of Head Start’s efforts to engage families in their children’s learning. Family engagement has been a priority in Head Start’s programming since its inception in 1965, a priority that remains strong today as evidenced by Head Start’s new
Parent, Family, and Community Engagement framework. In order to learn more about whether current family engagement efforts in Head Start align with the PFCE framework, a recent report highlighted patterns in the family engagement practices currently taking place in Head Start programs using data from the most recent wave of the Head Start Family and Child Experiences Survey (FACES 2014) (Aikens et al., 2017). This survey included a Family Engagement Plus study, which gathered descriptive quantitative and qualitative information about family engagement and service provision from Head Start parents, teachers, and family services staff. The vast majority of parents in the survey reported that Head Start program staff encourage them to engage in academic, learning, and language activities with their children and often provide suggestions of specific activities they can do at home to improve school readiness (Aikens et al., 2017). Most parents also reported that they communicate with their children’s teachers daily or nearly daily to discuss their children’s developmental needs and issues, and that these interactions typically happen during morning drop off or afternoon pickup. Family service staff similarly report having regular, sometimes daily, contact with families during informal pick up and drop off times, though they also report needing additional supports to more effectively engage families, such as additional funding and staff, reduced caseloads, and more opportunities for staff trainings. These results suggest that there is a great deal of mutual engagement on the parts of parents, teachers, and family service staff in Head Start programs today, but that additional resources and training could help to further bolster this engagement. Furthermore, although parents in this study acknowledged teachers’ and family service staff’s efforts to help them engage in their children’s learning at home, without a formal evaluation it remains unclear whether these efforts are truly effective at engaging parents in their children’s learning. Future
evaluations of Head Start should prioritize understanding the causal effect that these efforts have on parents’ engagement in and promotion of children’s learning.

**Head Start and Parents’ Harsh Discipline**

With regard to Head Start’s impact on parents’ disciplinary practices, I found that Head Start was minimally effective on average at lowering parents’ use and frequency of spanking. The effect for spanking frequency varied very slightly, though the magnitude of variation was very small and likely not practically meaningful. The lack of an average effect might be due to the sensitivity of parents’ discipline as a topic and the fact that lowering its use is not one of Head Start’s main goals, suggesting that Head Start centers may not prioritize reducing harsh discipline as a focus of intervention. Indeed, in the recent FACES 2014 Family Engagement Plus study, 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)—both of which imply that attempting to reduce parents’ use of harsh discipline is not a common activity for Head Start providers.

Across studies in the present dissertation, the only program or family characteristic to predict variation in Head Start effects on parents’ spanking use was parental immigrant status—centers serving higher proportions of immigrants were less effective at reducing parents’ use of spanking. It is not clear why this is, though the reason could have to do with greater overall use of spanking among Hispanic parents, who made up a large proportion of the immigrant families in the HSIS, compared to White parents (Gershoff et al., 2012). This proposition is highly speculative, however, and should be explored further before drawing firm conclusions.

Regardless of why centers serving relatively more immigrants were less effective at reducing spanking use, the general lack of Head Start effectiveness at reducing spanking is
notable. 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 disadvantaged parents—are problematic. The American Academy of Pediatrics recently 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, Siegel, & AAP Council on Child Abuse and Neglect, 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.

Remaining Questions and Future Research

Despite the important contributions of the three studies included in the present dissertation, there remain a number of important questions left to be answered about the potential for Head Start to promote positive parenting behaviors. First, I chose a number of theoretically relevant center and family characteristics that I hypothesized might predict which Head Start centers are more versus less effective. However, there are a number of other center and family characteristics that could play a role in this association. For instance, center characteristics such
as the racial and ethnic or language match between providers and parents might play a role in the effectiveness of provider-parent trust and communication (Barrueco et al., 2016). Thus it could be that centers whose providers and teachers are more often of the same race/ethnicity or speak the same language are more effective at promoting parenting outcomes than centers with less overlap in these domains. Similarly, family characteristics like parental stress and depression might also play a role in centers’ effectiveness. I also found that certain family characteristics such as household economic risk, MGHS status, and initial cognitive stimulation were positively correlated at the center level, meaning that centers that served high proportions of families with one of these characteristics were also likely to serve higher proportions of families with the other characteristics. It could be that the makeup of certain family characteristic “profiles” or combinations within Head Start centers contribute to centers’ effectiveness. For example, it could be that centers serving a high proportion of families with high household economic risk in which many parents attended Head Start and in which parents had high average levels of initial cognitive stimulation are more effective than centers with other demographic profiles. Factor analysis could reveal such “profiles” of Head Start centers that could be related to the effectiveness of centers. Similarly, it could be that it is centers with some combination of center and family characteristics (e.g. centers that provide frequent home visits and who also serve a high proportion of immigrant families) that are most effective. There are thus many more possibilities to be explored with regard to the potential for other center and family characteristics—or some combination of them—to predict variation in Head Start treatment effects on parenting outcomes.

There are also additional sources of variation left to explore. In their conceptual framework for studying the sources of variation in program effects, Weiss and colleagues (2014)
posited three factors that could theoretically moderate a program’s treatment effect on a given outcome: characteristics of the organization; characteristics of the individuals participating in the intervention; and the treatment contrast, or the difference between what the treatment and control group of an intervention experience. I was able to examine the role of some center and family characteristics in Head Start’s impact on parenting outcomes, but it was beyond the scope of this dissertation to also examine the role of the treatment contrast. In their investigation of whether Head Start effects on children’s outcomes varied depending on the counterfactual child care type (i.e. the setting in which children would have otherwise received care), Feller and colleagues (2016) used the HSIS with a principal stratification framework to define groups of children based on their observed care setting and their predicted counterfactual care setting. They found that it was children who complied with their random assignment to Head Start 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 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, given that Head Start places a larger emphasis on parent engagement than do other types of 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 impact 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.

Finally, while the HSIS is certainly the ideal dataset with which to explore questions
about Head Start impact variation, it is important to remember that the data are somewhat dated
and thus my findings with regard to the program and family characteristics that predict variation
in treatment effects may not be entirely generalizable to Head Start centers today. Head Start
may operate differently in 2019 than in 2002. Indeed, Head Start’s renewed commitment to
family engagement through the PFCE framework may be indicative of stronger parent outreach
services available to Head Start parents today that have important implications for Head Start’s
potentially greater ability to impact parenting outcomes. In addition to potentially greater overall
efforts to engage families today, Head Start centers are engaging families differently than they
were in 2002. In addition to using more traditional approaches to engaging families such as the
ones described in this dissertation, programs today also use technology such as text messaging,
email, and social media to engage families (Aikens et al., 2017). These changes have
implications for Head Start’s ability to reach parents and other family members, potentially
resulting in larger Head Start effects on promoting positive parenting behaviors today than were
observed in the HSIS. Furthermore, the national ECE landscape has changed dramatically since
2002, with many more states today providing universal public pre-k along with growing attention
to family engagement as a strategy for enhancing children’s outcomes (Center for the Study of
Social Policy, 2016). Thus, the treatment contrast between services for parents in Head Start
compared to other programs may be different today in ways relevant to Head Start’s effect on
parenting outcomes. Future research should investigate what publicly funded programs—both
Head Start and non-Head Start— are doing today to engage families and how this engagement impacts parents’ involvement in their children’s early learning.

Ultimately, understanding whether Head Start can promote positive parenting practices is an important endeavor because of the potential for improved parenting practices to improve child outcomes. It is thus also important to determine whether changes in parenting behaviors as a result of Head Start are large enough to result in changes in children’s outcomes. Future data collection efforts should therefore collect data from both parents and children at frequent time intervals so that longitudinal models can be developed to understand this dynamic process.

**Conclusions**

Parenting is the most potent influence on child outcomes, yet low-income parents on average provide fewer and lower quality developmentally-enriching interactions with children, and are more likely to engage in harsh, physical discipline, than their higher income counterparts. The present dissertation takes an important step towards understanding the full extent to which Head Start, a program that millions of low-income families engage with regularly, can serve as a naturally occurring parenting intervention. On average, Head Start has relatively small impacts on parents, but I found that some centers are much more, and some are much less effective than others. The program and family characteristics explored in this dissertation were somewhat predictive of differences in treatment effects across centers, but there are many questions left to be answered about how to make Head Start most effective for the families it serves. As one of the largest sources of federal support for low-income families in the United States, Head Start provides a unique and promising opportunity to enhance parent and therefore child outcomes. It is therefore important that we make the most of this opportunity, for it has the potential to meaningfully improve the lives of millions of children and families.
APPENDIX

List of individual items in the Fall Cognitive Stimulation composite

STEM: “In the past week, have you or someone in your family done the following things with [CHILD]?”

If the response was “yes,” parents were asked “How many times have you done this in the past week?”

ITEMS:

Told (him/her) a story
Helped (him/her) learn the names of letters, words, or numbers
Practiced writing the letters of the alphabet with (him/her)
Helped [CHILD] learn songs or music
Worked on arts or crafts with (him/her)
Played with toys or games indoors with [CHILD]
Played a sport or exercised together
Taken (him/her) along while doing errands
Involved (him/her) in household chores
Practiced writing or spelling [CHILD]’s name
Practiced rhyming words
Counted the number of things that you can see and touch
Talked about how big something is or how much something holds using a ruler, spoons, or cups;
Talked about the calendar or days of the week.

List of individual items in the Spring Literacy Activities composite

STEM: “During the day, how often do you or someone in your household do each of the following reading and language activities with [CHILD]?”

ITEMS:

Work on learning the names of the letters
Practice reading and writing the letters of the alphabet
Discuss new words
Have [CHILD] tell you a story
Practice the sounds that letters make or phonics
Listen to you read stories where (he/she) sees the print
Listen to you read stories where (he/she) doesn’t see the print
Retell or make up stories
Show [CHILD] how to read a book or magazine
Have [CHILD] practice writing or spelling (his/her) name
Learn about rhyming words and word families
Practice or teach directional words

**List of individual items in the Spring Math Activities composite**

STEM: “During the day, how often do you or someone in your household do each of the following math activities with [CHILD]?”

**ITEMS:**

- Count out loud
- Work or shape blocks
- Count things such as small toys or chips to learn math
- Play math-related games
- Use music to understand math ideas
- Use dance or act out stories to practice math ideas such as numbers, size, or shapes
- Work with rulers, measuring cups, spoons, or other measuring instruments
- Talk about the calendar or days of the week
### SUPPLEMENTARY TABLES

Table A1

*Models showing the non-causal ITT effect of ECERS-R subscales on the average Head Start effect*

<table>
<thead>
<tr>
<th></th>
<th>Child is read to daily (y/n)</th>
<th>p</th>
<th>Literacy Activities</th>
<th>p</th>
<th>Math activities</th>
<th>p</th>
<th>Child spanked in past week (y/n)</th>
<th>p</th>
<th>Number of spanks</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ECERS-R: Interactions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>0.033</td>
<td>0.426</td>
<td>0.075</td>
<td>0.407</td>
<td>0.071</td>
<td>0.427</td>
<td>-0.044</td>
<td>0.303</td>
<td>0.001</td>
<td>0.869</td>
</tr>
<tr>
<td>Treatment*Interactions</td>
<td>0.011</td>
<td>0.806</td>
<td>0.106</td>
<td>0.278</td>
<td>0.134</td>
<td>0.166</td>
<td>0.025</td>
<td>0.579</td>
<td>-0.004</td>
<td>0.682</td>
</tr>
<tr>
<td><strong>ECERS-R: Program Structure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>-0.001</td>
<td>0.984</td>
<td>0.115</td>
<td>0.109</td>
<td>0.198</td>
<td>0.006</td>
<td>**0.004</td>
<td>0.912</td>
<td>-0.002</td>
<td>0.744</td>
</tr>
<tr>
<td>Treatment*Program Structure</td>
<td>0.056</td>
<td>0.130</td>
<td>0.065</td>
<td>0.426</td>
<td>-0.017</td>
<td>0.834</td>
<td>-0.034</td>
<td>0.379</td>
<td>0.001</td>
<td>0.946</td>
</tr>
<tr>
<td><strong>ECERS-R: Language and reasoning</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>0.033</td>
<td>0.426</td>
<td>0.075</td>
<td>0.407</td>
<td>0.071</td>
<td>0.427</td>
<td>-0.044</td>
<td>0.303</td>
<td>0.001</td>
<td>0.869</td>
</tr>
<tr>
<td>Treatment*Language and reasoning</td>
<td>0.011</td>
<td>0.806</td>
<td>0.106</td>
<td>0.278</td>
<td>0.134</td>
<td>0.166</td>
<td>0.025</td>
<td>0.579</td>
<td>-0.004</td>
<td>0.682</td>
</tr>
<tr>
<td><strong>ECERS-R: Activities</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>0.04179</td>
<td>0.032</td>
<td>0.125</td>
<td>0.003</td>
<td><strong>0.195</strong></td>
<td>0.000</td>
<td><strong>-0.018</strong></td>
<td>0.377</td>
<td>-0.001</td>
<td>0.820</td>
</tr>
<tr>
<td>Treatment*Activities</td>
<td>0.00161</td>
<td>0.962</td>
<td>0.107</td>
<td>0.139</td>
<td>-0.028</td>
<td>0.703</td>
<td>-0.014</td>
<td>0.687</td>
<td>0.000</td>
<td>0.967</td>
</tr>
</tbody>
</table>

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 effect. Covariates include those used in the original HSIS 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, number of spanks) 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.

***p<.001, **p<.01, *p<.05, +p<.10
Table A2
Models showing the non-causal ITT effect of ECERS-R total scores on the average Head Start effect when "high" ECERS-R quality is defined as a score of six or above

<table>
<thead>
<tr>
<th></th>
<th>Child is read to daily (y/n)</th>
<th>p</th>
<th>Literacy Activities</th>
<th>p</th>
<th>Math activities</th>
<th>p</th>
<th>Child spanked in past week (y/n)</th>
<th>p</th>
<th>Number of spanks</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECERS-R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>0.045</td>
<td>0.007</td>
<td>**</td>
<td>0.155</td>
<td>***</td>
<td>0.000</td>
<td>***</td>
<td>0.184</td>
<td>0.000</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Treatment*High Quality</td>
<td>-0.029</td>
<td>0.613</td>
<td>0.070</td>
<td>0.563</td>
<td>0.019</td>
<td>0.875</td>
<td>0.051</td>
<td>0.387</td>
<td>-0.009</td>
</tr>
</tbody>
</table>

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 effect. Covariates include those used in the original HSIS 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, number of spanks) 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.

***p<.001, **p<.01, *p<.05, +p<.10
Table A3
Descriptive statistics showing child and family level center-level demographic characteristics stratified by high versus low family service worker outreach

<table>
<thead>
<tr>
<th></th>
<th>Overall mean/% (N=281)</th>
<th>Low outreach mean/% (N=141)</th>
<th>High outreach mean/% (N=140)</th>
<th>Difference</th>
<th>p-value of difference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Child characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child was in 3-year-old cohort (%)</td>
<td>0.592</td>
<td>0.611</td>
<td>0.573</td>
<td>-0.038</td>
<td>0.319</td>
</tr>
<tr>
<td>Female (%)</td>
<td>0.513</td>
<td>0.499</td>
<td>0.528</td>
<td>0.029</td>
<td>0.335</td>
</tr>
<tr>
<td>Black (%)</td>
<td>0.271</td>
<td>0.312</td>
<td>0.229</td>
<td>-0.083</td>
<td>0.073 *</td>
</tr>
<tr>
<td>Hispanic (%)</td>
<td>0.360</td>
<td>0.393</td>
<td>0.326</td>
<td>-0.067</td>
<td>0.170</td>
</tr>
<tr>
<td>White (%)</td>
<td>0.369</td>
<td>0.295</td>
<td>0.445</td>
<td>0.150</td>
<td>0.002 **</td>
</tr>
<tr>
<td>Home language is Spanish (%)</td>
<td>0.308</td>
<td>0.360</td>
<td>0.256</td>
<td>-0.104</td>
<td>0.024 *</td>
</tr>
<tr>
<td><strong>Family characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maternal age (years)</td>
<td>29.040</td>
<td>29.579</td>
<td>28.497</td>
<td>-1.081</td>
<td>0.018 *</td>
</tr>
<tr>
<td>Maternal education (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than HS</td>
<td>0.365</td>
<td>0.383</td>
<td>0.347</td>
<td>-0.036</td>
<td>0.245</td>
</tr>
<tr>
<td>HS</td>
<td>0.351</td>
<td>0.336</td>
<td>0.366</td>
<td>0.030</td>
<td>0.322</td>
</tr>
<tr>
<td>More than HS</td>
<td>0.284</td>
<td>0.281</td>
<td>0.286</td>
<td>0.005</td>
<td>0.850</td>
</tr>
<tr>
<td>Maternal marital status (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>0.381</td>
<td>0.371</td>
<td>0.390</td>
<td>0.018</td>
<td>0.602</td>
</tr>
<tr>
<td>Married</td>
<td>0.454</td>
<td>0.467</td>
<td>0.441</td>
<td>-0.027</td>
<td>0.458</td>
</tr>
<tr>
<td>Previously married</td>
<td>0.165</td>
<td>0.161</td>
<td>0.169</td>
<td>0.008</td>
<td>0.735</td>
</tr>
<tr>
<td>Mother is a teenager (%)</td>
<td>0.157</td>
<td>0.141</td>
<td>0.172</td>
<td>0.031</td>
<td>0.181</td>
</tr>
<tr>
<td>Mother is a recent immigrant (%)</td>
<td>0.188</td>
<td>0.210</td>
<td>0.165</td>
<td>-0.044</td>
<td>0.175</td>
</tr>
<tr>
<td>Child lives with both biological parents (%)</td>
<td>0.507</td>
<td>0.533</td>
<td>0.482</td>
<td>-0.051</td>
<td>0.155</td>
</tr>
<tr>
<td><strong>Assessment characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child age at spring testing (years)</td>
<td>4.472</td>
<td>4.443</td>
<td>4.502</td>
<td>0.059</td>
<td>0.188</td>
</tr>
<tr>
<td>Child was tested in Spanish (%)</td>
<td>0.261</td>
<td>0.294</td>
<td>0.227</td>
<td>-0.067</td>
<td>0.119</td>
</tr>
<tr>
<td>Time between baseline and spring parent interview (weeks)</td>
<td>33.703</td>
<td>33.786</td>
<td>33.618</td>
<td>-0.168</td>
<td>0.584</td>
</tr>
</tbody>
</table>

***p<.001, **p<.01, *p<.05, *p<.10
Table A4
Descriptive statistics showing center-level family demographic characteristics for centers differing in the percentage of moderate-high risk families they serve

<table>
<thead>
<tr>
<th></th>
<th>Overall mean/</th>
<th>Less than</th>
<th>40% and above</th>
<th>Difference</th>
<th>p-value of difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean/%</td>
<td>40% moderate-high risk families</td>
<td>moderate-high risk families</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Family Moderators</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of moderate-high risk families</td>
<td>0.254</td>
<td>0.134</td>
<td>0.567</td>
<td>0.433</td>
<td>&lt;0.0000 ***</td>
</tr>
<tr>
<td>Percentage of families with immigrant mothers</td>
<td>0.187</td>
<td>0.212</td>
<td>0.123</td>
<td>-0.089</td>
<td>0.0087 **</td>
</tr>
<tr>
<td>Percentage of multigenerational Head Start families</td>
<td>0.240</td>
<td>0.224</td>
<td>0.278</td>
<td>0.054</td>
<td>0.1015</td>
</tr>
<tr>
<td>Average initial cognitive stimulation score</td>
<td>1.156</td>
<td>1.143</td>
<td>1.190</td>
<td>0.047</td>
<td>0.1100</td>
</tr>
<tr>
<td>Percentage of families who initially spanked</td>
<td>0.416</td>
<td>0.385</td>
<td>0.495</td>
<td>0.110</td>
<td>0.0016 **</td>
</tr>
<tr>
<td><strong>Child characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child was in 3-year-old cohort (%)</td>
<td>0.596</td>
<td>0.594</td>
<td>0.604</td>
<td>0.011</td>
<td>0.7963</td>
</tr>
<tr>
<td>Female (%)</td>
<td>0.514</td>
<td>0.512</td>
<td>0.518</td>
<td>0.006</td>
<td>0.8511</td>
</tr>
<tr>
<td>Black (%)</td>
<td>0.275</td>
<td>0.239</td>
<td>0.370</td>
<td>0.131</td>
<td>0.0164 *</td>
</tr>
<tr>
<td>Hispanic (%)</td>
<td>0.362</td>
<td>0.389</td>
<td>0.289</td>
<td>-0.100</td>
<td>0.0527 +</td>
</tr>
<tr>
<td>White (%)</td>
<td>0.363</td>
<td>0.372</td>
<td>0.340</td>
<td>-0.031</td>
<td>0.5460</td>
</tr>
<tr>
<td>Home language is Spanish (%)</td>
<td>0.309</td>
<td>0.358</td>
<td>0.179</td>
<td>-0.179</td>
<td>0.0001 ***</td>
</tr>
<tr>
<td><strong>Family characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maternal age (years)</td>
<td>29.125</td>
<td>29.573</td>
<td>27.956</td>
<td>-1.616</td>
<td>0.0024 **</td>
</tr>
<tr>
<td>Maternal education (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than HS</td>
<td>0.366</td>
<td>0.340</td>
<td>0.436</td>
<td>0.096</td>
<td>0.0032 **</td>
</tr>
<tr>
<td>HS</td>
<td>0.351</td>
<td>0.362</td>
<td>0.322</td>
<td>-0.041</td>
<td>0.2048</td>
</tr>
<tr>
<td>More than HS</td>
<td>0.282</td>
<td>0.298</td>
<td>0.242</td>
<td>-0.055</td>
<td>0.0624 +</td>
</tr>
<tr>
<td>Maternal marital status (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>0.387</td>
<td>0.317</td>
<td>0.570</td>
<td>0.253</td>
<td>&lt;0.0000 ***</td>
</tr>
<tr>
<td>Married</td>
<td>0.451</td>
<td>0.522</td>
<td>0.266</td>
<td>-0.255</td>
<td>&lt;0.0000 ***</td>
</tr>
<tr>
<td>Previously married</td>
<td>0.162</td>
<td>0.161</td>
<td>0.164</td>
<td>0.003</td>
<td>0.9188</td>
</tr>
<tr>
<td>Mother is a teenager (%)</td>
<td>0.156</td>
<td>0.105</td>
<td>0.288</td>
<td>0.182</td>
<td>&lt;0.0000 ***</td>
</tr>
<tr>
<td>Mother is a recent immigrant (%)</td>
<td>0.187</td>
<td>0.212</td>
<td>0.123</td>
<td>-0.089</td>
<td>0.0087 **</td>
</tr>
<tr>
<td>Child lives with both biological parents (%)</td>
<td>0.505</td>
<td>0.563</td>
<td>0.352</td>
<td>-0.211</td>
<td>&lt;0.0000 ***</td>
</tr>
<tr>
<td><strong>Assessment characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child age at spring testing (years)</td>
<td>4.465</td>
<td>4.465</td>
<td>4.465</td>
<td>0.000</td>
<td>1.0000</td>
</tr>
<tr>
<td>Child was tested in Spanish (%)</td>
<td>0.257</td>
<td>0.288</td>
<td>0.177</td>
<td>-0.111</td>
<td>0.0098 **</td>
</tr>
<tr>
<td>Time between baseline and spring parent interview (weeks)</td>
<td>33.704</td>
<td>33.759</td>
<td>33.560</td>
<td>-0.199</td>
<td>0.5319</td>
</tr>
</tbody>
</table>

Notes: 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. Ns vary across variables because center-level data were not imputed and are thus missing for some centers. Percentage of families with moderate-high risk families is entered into all models continuously, but was dichotomized in this table for the purposes of showing how these family characteristics differ according to the percentage of moderate-high risk families served in a center. Above 40% moderate-high risk families was chosen to represent centers serving a "high" percentage of moderate-high risk families because this percentage represents the 75th percentile.

***p<.001, **p<.01, *p<.05, +p<.10
Table A5
Descriptive statistics showing center-level family demographic characteristics for centers differing in the percentage of immigrant families they serve

<table>
<thead>
<tr>
<th>Family Moderators</th>
<th>Overall mean/%</th>
<th>Less than 30% immigrants</th>
<th>30% and above immigrants</th>
<th>Difference</th>
<th>p-value of difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of moderate-high risk families</td>
<td>0.254</td>
<td>0.286</td>
<td>0.174</td>
<td>-0.112</td>
<td>0.0001 ***</td>
</tr>
<tr>
<td>Percentage of families with immigrant mothers</td>
<td>0.187</td>
<td>0.034</td>
<td>0.567</td>
<td>0.533</td>
<td>&lt;0.0000 ***</td>
</tr>
<tr>
<td>Percentage of multigenerational Head Start families</td>
<td>0.240</td>
<td>0.293</td>
<td>0.102</td>
<td>-0.191</td>
<td>&lt;0.0000 ***</td>
</tr>
<tr>
<td>Average initial cognitive stimulation score</td>
<td>1.156</td>
<td>1.198</td>
<td>1.052</td>
<td>-0.146</td>
<td>&lt;0.0000 ***</td>
</tr>
<tr>
<td>Percentage of families who initially spanked</td>
<td>0.416</td>
<td>0.433</td>
<td>0.372</td>
<td>-0.061</td>
<td>0.0737 +</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Child characteristics</th>
<th>Overall mean/%</th>
<th>Less than 30% immigrants</th>
<th>30% and above immigrants</th>
<th>Difference</th>
<th>p-value of difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child was in 3-year-old cohort (%)</td>
<td>0.596</td>
<td>0.643</td>
<td>0.481</td>
<td>-0.162</td>
<td>0.0003 ***</td>
</tr>
<tr>
<td>Female (%)</td>
<td>0.514</td>
<td>0.519</td>
<td>0.501</td>
<td>-0.018</td>
<td>0.5814</td>
</tr>
<tr>
<td>Black (%)</td>
<td>0.275</td>
<td>0.356</td>
<td>0.074</td>
<td>-0.282</td>
<td>&lt;0.0000 ***</td>
</tr>
<tr>
<td>Hispanic (%)</td>
<td>0.362</td>
<td>0.190</td>
<td>0.789</td>
<td>0.599</td>
<td>&lt;0.0000 ***</td>
</tr>
<tr>
<td>White (%)</td>
<td>0.363</td>
<td>0.454</td>
<td>0.137</td>
<td>-0.317</td>
<td>&lt;0.0000 ***</td>
</tr>
<tr>
<td>Home language is Spanish (%)</td>
<td>0.309</td>
<td>0.125</td>
<td>0.764</td>
<td>0.640</td>
<td>&lt;0.0000 ***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Family characteristics</th>
<th>Overall mean/%</th>
<th>Less than 30% immigrants</th>
<th>30% and above immigrants</th>
<th>Difference</th>
<th>p-value of difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maternal age (years)</td>
<td>29.125</td>
<td>28.729</td>
<td>30.108</td>
<td>1.378</td>
<td>0.0055 **</td>
</tr>
<tr>
<td>Maternal education (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0000</td>
</tr>
<tr>
<td>Less than HS</td>
<td>0.366</td>
<td>0.315</td>
<td>0.495</td>
<td>0.180</td>
<td>&lt;0.0000 ***</td>
</tr>
<tr>
<td>HS</td>
<td>0.351</td>
<td>0.390</td>
<td>0.255</td>
<td>-0.135</td>
<td>&lt;0.0000 ***</td>
</tr>
<tr>
<td>More than HS</td>
<td>0.282</td>
<td>0.295</td>
<td>0.250</td>
<td>-0.046</td>
<td>0.1388</td>
</tr>
<tr>
<td>Maternal marital status (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0000</td>
</tr>
<tr>
<td>Single</td>
<td>0.387</td>
<td>0.419</td>
<td>0.309</td>
<td>-0.110</td>
<td>0.0017 **</td>
</tr>
<tr>
<td>Married</td>
<td>0.451</td>
<td>0.407</td>
<td>0.559</td>
<td>0.152</td>
<td>&lt;0.0000 ***</td>
</tr>
<tr>
<td>Previously married</td>
<td>0.162</td>
<td>0.174</td>
<td>0.131</td>
<td>-0.043</td>
<td>0.0795 +</td>
</tr>
<tr>
<td>Mother is a teenager (%)</td>
<td>0.156</td>
<td>0.176</td>
<td>0.107</td>
<td>-0.068</td>
<td>0.0033 **</td>
</tr>
<tr>
<td>Mother is a recent immigrant (%)</td>
<td>0.187</td>
<td>0.034</td>
<td>0.567</td>
<td>0.533</td>
<td>&lt;0.0000 ***</td>
</tr>
<tr>
<td>Child lives with both biological parents (%)</td>
<td>0.505</td>
<td>0.451</td>
<td>0.638</td>
<td>0.187</td>
<td>&lt;0.0000 ***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Assessment characteristics</th>
<th>Overall mean/%</th>
<th>Less than 30% immigrants</th>
<th>30% and above immigrants</th>
<th>Difference</th>
<th>p-value of difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child age at spring testing (years)</td>
<td>4.465</td>
<td>4.430</td>
<td>4.553</td>
<td>0.123</td>
<td>0.0173 *</td>
</tr>
<tr>
<td>Child was tested in Spanish (%)</td>
<td>0.257</td>
<td>0.089</td>
<td>0.674</td>
<td>0.584</td>
<td>&lt;0.0000 ***</td>
</tr>
<tr>
<td>Time between baseline and spring parent interview (weeks)</td>
<td>33.704</td>
<td>33.503</td>
<td>34.202</td>
<td>0.699</td>
<td>0.0425 *</td>
</tr>
</tbody>
</table>

Notes: 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. Ns vary across variables because center-level data were not imputed and are thus missing for some centers.
Percentage of families with immigrant mothers is entered into all models continuously, but was dichotomized in this table for the purposes of showing how these family characteristics differ according to the percentage of immigrants served in a center. Above 30% immigrants was chosen to represent centers serving a "high" percentage of immigrants because this percentage roughly represents the 75th percentile.

***p<.001, **p<.01, *p<.05, +p<.10
Table A6
Descriptive statistics showing center-level family demographic characteristics for centers differing in the percentage of multigenerational Head Start (MGHS) families they serve

<table>
<thead>
<tr>
<th>Family Moderators</th>
<th>Overall mean/</th>
<th>Less than 40% MGHS families</th>
<th>40% and above MGHS families</th>
<th>Difference</th>
<th>p-value of difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of moderate-high risk families</td>
<td>0.260</td>
<td>0.237</td>
<td>0.320</td>
<td>0.083</td>
<td>0.0065 **</td>
</tr>
<tr>
<td>Percentage of families with immigrant mothers</td>
<td>0.183</td>
<td>0.227</td>
<td>0.064</td>
<td>-0.164</td>
<td>&lt;0.0000 ***</td>
</tr>
<tr>
<td>Percentage of multigenerational Head Start families</td>
<td>0.240</td>
<td>0.107</td>
<td>0.601</td>
<td>0.494</td>
<td>&lt;0.0000 ***</td>
</tr>
<tr>
<td>Average initial cognitive stimulation score</td>
<td>1.158</td>
<td>1.122</td>
<td>1.255</td>
<td>0.133</td>
<td>&lt;0.0000 ***</td>
</tr>
<tr>
<td>Percentage of families who initially spanked</td>
<td>0.420</td>
<td>0.416</td>
<td>0.431</td>
<td>0.015</td>
<td>0.666</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Child characteristics</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Child was in 3-year-old cohort (%)</td>
<td>0.589</td>
<td>0.567</td>
<td>0.648</td>
<td>0.080</td>
<td>0.0509 +</td>
</tr>
<tr>
<td>Female (%)</td>
<td>0.509</td>
<td>0.505</td>
<td>0.519</td>
<td>0.015</td>
<td>0.6568</td>
</tr>
<tr>
<td>Black (%)</td>
<td>0.283</td>
<td>0.197</td>
<td>0.516</td>
<td>0.319</td>
<td>&lt;0.0000 ***</td>
</tr>
<tr>
<td>Hispanic (%)</td>
<td>0.351</td>
<td>0.416</td>
<td>0.174</td>
<td>-0.242</td>
<td>&lt;0.0000 ***</td>
</tr>
<tr>
<td>White (%)</td>
<td>0.366</td>
<td>0.387</td>
<td>0.310</td>
<td>-0.077</td>
<td>0.1495</td>
</tr>
<tr>
<td>Home language is Spanish (%)</td>
<td>0.295</td>
<td>0.363</td>
<td>0.110</td>
<td>-0.253</td>
<td>&lt;0.0000 ***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Family characteristics</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Maternal age (years)</td>
<td>28.888</td>
<td>29.293</td>
<td>27.782</td>
<td>-1.511</td>
<td>0.0011 **</td>
</tr>
<tr>
<td>Maternal education (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than HS</td>
<td>0.366</td>
<td>0.382</td>
<td>0.320</td>
<td>-0.062</td>
<td>0.0492 *</td>
</tr>
<tr>
<td>HS</td>
<td>0.354</td>
<td>0.336</td>
<td>0.404</td>
<td>0.067</td>
<td>0.0626 +</td>
</tr>
<tr>
<td>More than HS</td>
<td>0.280</td>
<td>0.282</td>
<td>0.276</td>
<td>-0.006</td>
<td>0.8591</td>
</tr>
<tr>
<td>Maternal marital status (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>0.394</td>
<td>0.345</td>
<td>0.528</td>
<td>0.183</td>
<td>&lt;0.0000 ***</td>
</tr>
<tr>
<td>Married</td>
<td>0.441</td>
<td>0.496</td>
<td>0.291</td>
<td>-0.206</td>
<td>&lt;0.0000 ***</td>
</tr>
<tr>
<td>Previously married</td>
<td>0.165</td>
<td>0.159</td>
<td>0.182</td>
<td>0.023</td>
<td>0.4254</td>
</tr>
<tr>
<td>Mother is a teenager (%)</td>
<td>0.158</td>
<td>0.143</td>
<td>0.198</td>
<td>0.055</td>
<td>0.0449 *</td>
</tr>
<tr>
<td>Mother is a recent immigrant (%)</td>
<td>0.183</td>
<td>0.227</td>
<td>0.064</td>
<td>-0.164</td>
<td>&lt;0.0000 ***</td>
</tr>
<tr>
<td>Child lives with both biological parents (%)</td>
<td>0.501</td>
<td>0.552</td>
<td>0.363</td>
<td>-0.190</td>
<td>&lt;0.0000 ***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Assessment characteristics</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Child age at spring testing (years)</td>
<td>4.471</td>
<td>4.483</td>
<td>4.436</td>
<td>-0.047</td>
<td>0.3325</td>
</tr>
<tr>
<td>Child was tested in Spanish (%)</td>
<td>0.246</td>
<td>0.308</td>
<td>0.076</td>
<td>-0.232</td>
<td>&lt;0.0000 ***</td>
</tr>
<tr>
<td>Time between baseline and spring parent interview (weeks)</td>
<td>33.713</td>
<td>33.628</td>
<td>33.945</td>
<td>0.317</td>
<td>0.3811</td>
</tr>
</tbody>
</table>

Notes: 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. Ns vary across variables because center-level data were not imputed and are thus missing for some centers. Percentage of MGHS families is entered into all models continuously, but was dichotomized in this table for the purposes of showing how these family characteristics differ according to the percentage of MGHS families served in a center. Above 40% MGHS families was chosen to represent centers serving a "high" percentage of MGHS families because this percentage roughly represents the 75th percentile.

***p<.001, **p<.01, *p<.05, +p<.10
Table A7
Descriptive statistics showing center-level family demographic characteristics for centers differing in the average initial cognitive stimulation among families they serve

<table>
<thead>
<tr>
<th></th>
<th>Overall mean/%</th>
<th>Top 25% average pre-academic stimulation</th>
<th>Bottom 75% average pre-academic stimulation</th>
<th>Difference</th>
<th>p-value of difference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Family Moderators</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of moderate-high risk families</td>
<td>0.255</td>
<td>0.236</td>
<td>0.307</td>
<td>0.071</td>
<td>0.0402 *</td>
</tr>
<tr>
<td>Percentage of families with immigrant mothers</td>
<td>0.184</td>
<td>0.214</td>
<td>0.101</td>
<td>-0.114</td>
<td>0.0004 ***</td>
</tr>
<tr>
<td>Percentage of multigenerational Head Start families</td>
<td>0.240</td>
<td>0.200</td>
<td>0.349</td>
<td>0.148</td>
<td>0.0002 ***</td>
</tr>
<tr>
<td>Average initial cognitive stimulation score</td>
<td>1.156</td>
<td>1.060</td>
<td>1.430</td>
<td>0.370</td>
<td>&lt;0.0000 ***</td>
</tr>
<tr>
<td>Percentage of families who initially spanked</td>
<td>0.416</td>
<td>0.424</td>
<td>0.391</td>
<td>-0.033</td>
<td>0.3784</td>
</tr>
<tr>
<td><strong>Child characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child was in 3-year-old cohort (%)</td>
<td>0.595</td>
<td>0.604</td>
<td>0.571</td>
<td>-0.033</td>
<td>0.4519</td>
</tr>
<tr>
<td>Female (%)</td>
<td>0.512</td>
<td>0.498</td>
<td>0.552</td>
<td>0.055</td>
<td>0.1028</td>
</tr>
<tr>
<td>Black (%)</td>
<td>0.276</td>
<td>0.236</td>
<td>0.391</td>
<td>0.154</td>
<td>0.0068 **</td>
</tr>
<tr>
<td>Hispanic (%)</td>
<td>0.359</td>
<td>0.404</td>
<td>0.234</td>
<td>-0.170</td>
<td>0.0007 ***</td>
</tr>
<tr>
<td>White (%)</td>
<td>0.364</td>
<td>0.360</td>
<td>0.375</td>
<td>0.015</td>
<td>0.7872</td>
</tr>
<tr>
<td>Home language is Spanish (%)</td>
<td>0.306</td>
<td>0.351</td>
<td>0.178</td>
<td>-0.173</td>
<td>0.0003 ***</td>
</tr>
<tr>
<td><strong>Family characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maternal age (years)</td>
<td>29.061</td>
<td>29.089</td>
<td>28.981</td>
<td>-0.109</td>
<td>0.8316</td>
</tr>
<tr>
<td>Maternal education (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than HS</td>
<td>0.368</td>
<td>0.379</td>
<td>0.336</td>
<td>-0.043</td>
<td>0.1712</td>
</tr>
<tr>
<td>HS</td>
<td>0.352</td>
<td>0.334</td>
<td>0.405</td>
<td>0.071</td>
<td>0.0425 *</td>
</tr>
<tr>
<td>More than HS</td>
<td>0.280</td>
<td>0.287</td>
<td>0.259</td>
<td>-0.028</td>
<td>0.4008</td>
</tr>
<tr>
<td>Maternal marital status (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>0.389</td>
<td>0.369</td>
<td>0.444</td>
<td>0.074</td>
<td>0.0733 +</td>
</tr>
<tr>
<td>Married</td>
<td>0.449</td>
<td>0.477</td>
<td>0.371</td>
<td>-0.105</td>
<td>0.0078 **</td>
</tr>
<tr>
<td>Previously married</td>
<td>0.162</td>
<td>0.154</td>
<td>0.185</td>
<td>0.031</td>
<td>0.2474</td>
</tr>
<tr>
<td>Mother is a teenager (%)</td>
<td>0.157</td>
<td>0.146</td>
<td>0.187</td>
<td>0.041</td>
<td>0.1427</td>
</tr>
<tr>
<td>Mother is a recent immigrant (%)</td>
<td>0.184</td>
<td>0.214</td>
<td>0.101</td>
<td>-0.114</td>
<td>0.0004 ***</td>
</tr>
<tr>
<td>Child lives with both biological parents (%)</td>
<td>0.503</td>
<td>0.531</td>
<td>0.423</td>
<td>-0.108</td>
<td>0.0046 **</td>
</tr>
<tr>
<td><strong>Assessment characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child age at spring testing (years)</td>
<td>4.467</td>
<td>4.451</td>
<td>4.512</td>
<td>0.061</td>
<td>0.2077</td>
</tr>
<tr>
<td>Child was tested in Spanish (%)</td>
<td>0.255</td>
<td>0.293</td>
<td>0.146</td>
<td>-0.147</td>
<td>0.0006 ***</td>
</tr>
<tr>
<td>Time between baseline and spring parent interview (weeks)</td>
<td>33.693</td>
<td>33.821</td>
<td>33.329</td>
<td>-0.492</td>
<td>0.1400</td>
</tr>
</tbody>
</table>

Notes: 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. Ns vary across variables because center-level data were not imputed and are thus missing for some centers. Average pre-academic stimulation is entered into all models continuously, but was dichotomized in this table for the purposes of showing how these family characteristics differ according to centers’ average pre-academic stimulation. The top 25% of centers represent those with average pre-academic stimulation scores of 6.2 or above; the bottom 75% represent those with average scores of 6.1 or below. ***p<.001, **p<.01, *p<.05, *p<.10
Table A8
Descriptive statistics showing center-level family demographic characteristics for centers differing in the percentage of parents who reported initially spanking they serve

<table>
<thead>
<tr>
<th>Family Moderators</th>
<th>Overall mean/%</th>
<th>Less than 60% of parents initially spanked</th>
<th>More than 60% of parents initially spanked</th>
<th>Difference</th>
<th>p-value of difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of moderate-high risk families</td>
<td>0.255</td>
<td>0.244</td>
<td>0.294</td>
<td>0.050</td>
<td>0.1712</td>
</tr>
<tr>
<td>Percentage of families with immigrant mothers</td>
<td>0.184</td>
<td>0.198</td>
<td>0.138</td>
<td>-0.060</td>
<td>0.1013</td>
</tr>
<tr>
<td>Percentage of multigenerational Head Start families</td>
<td>0.240</td>
<td>0.246</td>
<td>0.215</td>
<td>-0.031</td>
<td>0.3749</td>
</tr>
<tr>
<td>Average initial cognitive stimulation score</td>
<td>1.156</td>
<td>1.164</td>
<td>1.130</td>
<td>-0.034</td>
<td>0.3708</td>
</tr>
<tr>
<td>Percentage of families who initially spanked</td>
<td>0.416</td>
<td>0.310</td>
<td>0.792</td>
<td>0.482</td>
<td>&lt;0.0000 ***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Child characteristics</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Child was in 3-year-old cohort (%)</td>
<td>0.595</td>
<td>0.580</td>
<td>0.648</td>
<td>0.067</td>
<td>0.1027</td>
</tr>
<tr>
<td>Female (%)</td>
<td>0.512</td>
<td>0.499</td>
<td>0.558</td>
<td>0.059</td>
<td>0.1288</td>
</tr>
<tr>
<td>Black (%)</td>
<td>0.276</td>
<td>0.252</td>
<td>0.364</td>
<td>0.112</td>
<td>0.0547 +</td>
</tr>
<tr>
<td>Hispanic (%)</td>
<td>0.359</td>
<td>0.389</td>
<td>0.254</td>
<td>-0.135</td>
<td>0.0135 *</td>
</tr>
<tr>
<td>White (%)</td>
<td>0.364</td>
<td>0.359</td>
<td>0.382</td>
<td>0.022</td>
<td>0.6968</td>
</tr>
<tr>
<td>Home language is Spanish (%)</td>
<td>0.306</td>
<td>0.339</td>
<td>0.191</td>
<td>-0.148</td>
<td>0.0027 **</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Family characteristics</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Maternal age (years)</td>
<td>29.061</td>
<td>29.109</td>
<td>28.893</td>
<td>-0.216</td>
<td>0.7000</td>
</tr>
<tr>
<td>Maternal education (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than HS</td>
<td>0.368</td>
<td>0.368</td>
<td>0.365</td>
<td>-0.003</td>
<td>0.9401</td>
</tr>
<tr>
<td>HS</td>
<td>0.352</td>
<td>0.355</td>
<td>0.342</td>
<td>-0.013</td>
<td>0.7058</td>
</tr>
<tr>
<td>More than HS</td>
<td>0.280</td>
<td>0.276</td>
<td>0.292</td>
<td>0.016</td>
<td>0.6595</td>
</tr>
<tr>
<td>Maternal marital status (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>0.389</td>
<td>0.373</td>
<td>0.444</td>
<td>0.071</td>
<td>0.1184</td>
</tr>
<tr>
<td>Married</td>
<td>0.449</td>
<td>0.465</td>
<td>0.391</td>
<td>-0.074</td>
<td>0.0923 +</td>
</tr>
<tr>
<td>Previously married</td>
<td>0.162</td>
<td>0.162</td>
<td>0.164</td>
<td>0.003</td>
<td>0.9226</td>
</tr>
<tr>
<td>Mother is a teenager (%)</td>
<td>0.157</td>
<td>0.156</td>
<td>0.158</td>
<td>0.002</td>
<td>0.9352</td>
</tr>
<tr>
<td>Mother is a recent immigrant (%)</td>
<td>0.184</td>
<td>0.198</td>
<td>0.138</td>
<td>-0.060</td>
<td>0.1013</td>
</tr>
<tr>
<td>Child lives with both biological parents (%)</td>
<td>0.503</td>
<td>0.519</td>
<td>0.447</td>
<td>-0.072</td>
<td>0.0873 +</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Assessment characteristics</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Child age at spring testing (years)</td>
<td>4.467</td>
<td>4.481</td>
<td>4.415</td>
<td>-0.066</td>
<td>0.1705</td>
</tr>
<tr>
<td>Child was tested in Spanish (%)</td>
<td>0.255</td>
<td>0.283</td>
<td>0.155</td>
<td>-0.128</td>
<td>0.0040 **</td>
</tr>
<tr>
<td>Time between baseline and spring parent interview (weeks)</td>
<td>33.693</td>
<td>33.875</td>
<td>33.049</td>
<td>-0.826</td>
<td>0.0137 *</td>
</tr>
</tbody>
</table>

Notes: 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. Ns vary across variables because center-level data were not imputed and are thus missing for some centers.
Percentage of families who initially spanked is entered into all models continuously, but was dichotomized in this table for the purposes of showing how these family characteristics differ according to the percentage of families who spanked within a center. Above 60% of parents who spanked was chosen to represent centers serving a "high" percentage of families who spanked because this percentage roughly represents the 75th percentile.

***p<.001, **p<.01, *p<.05, *p<.10
REFERENCES


https://doi.org/10.3368/jhr.43.1.139


program in early childhood, 1–49.


160


https://doi.org/10.1257/app.1.3.111


Easterbrooks, A. M., Dym Bartlett, J., Raskin, M., Goldberg, J., Contreras, M. M., Kotake, C.,


169


Paschall, K. W., & Mastergeorge, A. M. (2016). A review of 25 years of research in


Prinz, R. J., & Sanders, M. R. (2007). Adopting a population-level approach to parenting and
family support interventions, 27, 739–749. https://doi.org/10.1016/j.cpr.2007.01.005


Reardon, S. F., & Stuart, E. A. (2017). Editors’ introduction: Theme issue on variation in
https://doi.org/10.1080/19345747.2017.1386037


https://doi.org/10.1177/0956797612466268

https://doi.org/10.1016/j.chiabu.2010.03.004


Teitler, J. O., Reichman, N. E., & Nepomnyaschy, L. (2004). Sources of support, child care, and


https://doi.org/10.1111/j.1741-3737.2001.00136.x


https://doi.org/10.1016/j.childyouth.2011.03.008


Avenue, New York, NY 10022 ($22.50). Retrieved from https://eric.ed.gov/?id=ED183266