Head Start Impact Study
First Year Findings

June 2005

U.S. Department of Health and Human Services
Administration for Children and Families
Office of Planning, Research and Evaluation
Administration on Children, Youth and Families
Head Start Bureau

Head Start Impact Study

First Year Findings

June 2005

Prepared for:

Office of Planning, Research and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services
Washington, D.C.
under contract 282-00-0022, Head Start Impact Study

Prepared by:

Westat
1650 Research Blvd.
Rockville, MD 20850

The Urban Institute
2100 M Street, N.W.
Washington, DC 20037

Chesapeake Research Associates
708 Riverview Terrace
Annapolis, MD 21401

Decision Information Resources, Inc.
2600 Southwest Freeway, Suite 900
Houston, Texas 77098

American Institutes for Research
1000 Thomas Jefferson Street, N.W.
Washington, DC 20007

Head Start Impact Study: First Year Findings

Authors

Prepared by:

Michael Puma
Stephen Bell
Ronna Cook
Camilla Heid
Michael Lopez

Contributing Authors:

Nicholas Zill
Gary Shapiro
Pam Broene
Debra Mekos
Monica Rohacek
Liz Quinn
Gina Adams
Janet Friedman
Haidee Bernstein

Suggested Citation:

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

Acknowledgements

This report of the Head Start Impact Study focuses on the preliminary findings after the first Head Start program year. To get to this point required the hard work, dedication and partnership of people from a variety of venues. The expert leadership of Dr. Michael Lopez, the Federal Project Officer, was instrumental in making this project a reality. His vision, tenacity, and knowledge from the inception of the project to the completion of this report were remarkable. Our thanks are also extended to Jonathan Miles, whose tenure as an SRCD fellow at ACF provided substantial input into the analytic challenges of the study.

There were those who were worried that random assignment and subsequent data collection efforts would be difficult, if not impossible to implement. Study staff have done a tremendous job in meeting these challenges to ensure the success of the study. Moreover, the partnership and support from the National Head Start Association, Head Start Grantees and Delegate Agencies and their center staff were instrumental in the successful implementation of this study. The ongoing backing of the Head Start Bureau and Regional Office staff were critical to the recruitment process. A special thank you is extended to all the families and their children who participated in the study. Their continued contributions of time and information over the last three years have been exceptional and greatly appreciated.

We also want to thank the many external experts who helped us along the way, particularly the members of the Advisory Committee on Head Start Research and Evaluation. Your wisdom about sample design, measures, program, policy, and analytic challenges has helped formulate this product.

Finally, we gratefully acknowledge the staff from Westat, Chesapeake Research Associates, Urban Institute, American Institutes for Research, and Decision Information Resources, Inc. for their hard work, professionalism and dedication to the project.

* *

This report presents initial findings from the Congressionally mandated Head Start Impact Study. Three required reports to Congress have already been submitted. This report, while not mandated, presents preliminary findings on impacts after one year in Head Start (fall 2002 to spring 2003). A final report will present results of analyses following children through the end of first grade.

TABLE OF CONTENTS

List of Appendices

List of Exhibits

List of Exhibits

List of Exhibits

Executive Summary

Highlights

The Congressionally-mandated Head Start Impact Study is being conducted across 84 nationally representative grantee/delegate agencies. Approximately 5,000 newly entering 3- and 4-year-old children applying for Head Start were randomly assigned to either a Head Start group that had access to Head Start program services or to a non-Head Start group that could enroll in available community non-Head Start services, selected by their parents. Data collection began in fall 2002 and is scheduled to continue through 2006, following children through the spring of their 1st-grade year.

The study quantifies the impact of Head Start separately for 3- and 4-year-old children across child cognitive, social-emotional, and health domains as well as

Exhibit 1: Summary of Main Impact Findings[1]

Domains, Constructs, and Measures	Effect Sizes[2]	
	3-Year-Old Group	4-Year-Old Group
Cognitive Domain		
Pre-Reading		
Woodcock-Johnson III Letter-Word Identification	**0.24**	**0.22**
Letter Naming	**0.19**	**0.24**
Pre-Writing		
McCarthy Draw-A-Design	**0.13**	**--**
Woodcock-Johnson III Spelling	**--**	**0.16**
Vocabulary		
PPVT-III Adapted	**0.12**	**--**
Color Naming	**0.10**	--
Parent Reported Literacy Skills	**0.34**	**0.29**
Oral Comprehension and Phonological Awareness	**--**	**--**
Early Math	**--**	**--**
Social-Emotional Domain		
Problem Behaviors		
Total Behavior Problems	**-0.13**[3]	**--**
Hyperactive Behavior	**-0.18**[3]	**--**
Aggressive Behavior	--	--
Withdrawn Behavior	--	--
Social Skills and Approaches to Learning	**--**	**--**
Social Competencies	**--**	**--**
Health Domain		
Access to Health Care		
Child Had Dental Care	**0.34**	**0.32**
Child Has Health Insurance	--	--
Health Status		
Overall Health Status	**0.12**	**--**
Child Needs Ongoing Care	**--**	**--**
Child Had Care for Injury	**--**	**--**
Parenting Domain		
Educational Activities		
Number of Times Child Read To	**0.18**	**0.13**
Family Cultural Enrichment Scale	**0.11**	**--**
Discipline Strategies		
Spanked Child in Last Week	**-0.14**[3]	**--**
Number of Times Spanked	**-0.10**[3]	**--**
Used Timeout	**--**	**--**
Number of Timeouts	**--**	**--**
Child Safety Practices		
Overall Parental Safety Practices	**--**	**--**
Removing Harmful Objects	**--**	**--**
Restricting Child Movement	**--**	**--**
Safety Devices	**--**	**--**

[1] All effect sizes presented in table are based on statistically significant treatment and control differences of at least $p \leq 0.05$.
[2] Effect sizes relate the magnitude of impacts to the variation of the outcome as measured by the estimated treatment and control differences relative to the magnitude of the standard deviation on the measure of interest (i.e., as a fraction of one standard deviation).
[3] Negative effect sizes mean reduction in total problem behaviors, hyperactive behavior, and spanking.

on parenting practices. For children in the 3-year-old group, the preliminary results from the first year of data collection demonstrate small to moderate[1] positive effects favoring the children enrolled in Head Start for some outcomes in each domain. Fewer positive impacts were found for children in the 4-year-old group.[2] The key findings are summarized below and presented in Exhibit 1:

Cognitive Domain

The cognitive domain consists of six constructs each comprising one or more measures. The key findings in this domain are:

- There are small to moderate statistically significant positive impacts for both 3- and 4-year-old children on several measures across four of the six cognitive constructs, including pre-reading, pre-writing, vocabulary, and parent reports of children's literacy skills.
- No significant impacts were found for the constructs oral comprehension and phonological awareness or early mathematics skills for either age group.

Social-Emotional Domain

The social-emotional domain consists of three constructs, each comprising one or more parent-reported measures.[3] The key findings in this domain are:

- For children who entered the study as 3-year-olds, there is a small statistically significant impact in one of the three social-emotional constructs, problem behaviors.
- There were no statistically significant impacts on social skills and approaches to learning or on social competencies for 3-year-olds.
- No significant impacts were found for children entering the program as 4-year-olds.

Health Domain

The key findings in this domain, consisting of two constructs, are:

- For 3-year-olds, there are small to moderate statistically significant impacts in both constructs, higher parent reports of children's access to health care and reportedly better health status for children enrolled in Head Start.
- For children who entered the program as 4-year-olds, there are moderate statistically significant impacts on access to health care, but no significant impacts for health status.

---

[1] For this report we have adopted the following conventions for interpreting effect sizes: less than 0.2 is small, between 0.2 and 0.5 is a moderate impact, and over 0.5 is a large impact.

[2] Future analysis will test statistical significance of the differences in impacts across the two age groups.

[3] Future reports will also examine this domain using teacher-reported data.

Parenting Practices Domain

The key findings in this domain, consisting of three constructs, are:

- For children who entered the program as 3-year-olds, there are small statistically significant impacts in two of the three parenting constructs, including a higher use of educational activities and a lower use of physical discipline by parents of Head Start children. There were no significant impacts for safety practices.
- For children who entered the program as 4-year-olds, there are small statistically significant impacts on parents' use of educational activities. No significant impacts were found for discipline or safety practices.

Future reports will extend these analyses to examine additional areas of possible impact, explore possible variation in impact by program characteristics (e.g., classroom quality, teacher educational level, full-day versus part-day programs, etc.) and community characteristics, and follow children through the end of 1st grade.

Study Overview

Since its beginning in 1965 as a part of the War on Poverty, Head Start's goal has been to boost the school readiness of low-income children. Based on a "whole child" model, the program provides comprehensive services that include preschool education; medical, dental, and mental health care; nutrition services; and efforts to help parents foster their child's development. Head Start services are designed to be responsive to each child's and family's ethnic, cultural, and linguistic heritage.

Study Goals

1) Determine the impact of Head Start on:
   - Children's school readiness, and
   - Parental practices that support children's development.

2) Determine under what circumstances Head Start achieves its greatest impact and for which children.

In the 1998 reauthorization of Head Start, Congress mandated that the US Department of Health and Human Services (DHHS) determine, on a national level, the impact of Head Start on the children it serves. As noted by the Advisory Committee on Head Start Research, this legislative mandate required that the impact study address two main research questions:[4]

- "What difference does Head Start make to key outcomes of development and learning (and in particular, the multiple domains of school readiness) for low-income

[4]Advisory Committee on Head Start Research and Evaluation (1999). ***Evaluating Head Start: A Recommended Framework for Studying the Impact of the Head Start Program.*** Washington, DC: US Department of Health and Human Services.

children? What difference does Head Start make to parental practices that contribute to children's school readiness?"

- "Under what circumstances does Head Start achieve the greatest impact? What works for which children? What Head Start services are most related to impact?"

Random Assignment
Newly entering 3- and 4-year-old Head Start applicants were **randomly assigned** either to a **treatment group** that had access to Head Start services or to a **control group** that could receive any other non-Head Start services chosen by their parents.

To reliably answer these questions, a nationally representative sample of Head Start programs and newly entering 3- and 4-year-old children was selected, and children were **randomly assigned** either to a **treatment group** that had access to Head Start services or to a **control group** that could receive any other non-Head Start services available in the community, chosen by their parents. Under this randomized design, a simple comparison of outcomes for the two groups yields an unbiased estimate of the impact of access to Head Start on children's school readiness. This research design, if properly implemented, ensures that the two groups will not differ in any systematic or unmeasured way except through their access to Head Start services.

In addition to random assignment, this study is set apart from most program evaluations because children were selected at random from those applying for entry into Head Start in a nationally representative sample of programs, making results generalizable to the entire Head Start program, not just to the selected samples of programs and children.

Study Sample
The nationally representative study sample, spread over 23 different states, consists of a total of 84 randomly selected grantees/delegate agencies, 383 randomly selected Head Start centers, and a total of 4,667 newly entering children; 2,559 3-year-olds and 2,108 4-year-olds.

One constraint imposed on this study was that selected Head Start grantees and centers had to have a sufficient number of "extra" applicants for the 2002-03 program year to allow for the creation of a non-Head Start control group through random assignment, thereby avoiding ethical concerns about possible denial of services to eligible children. As a consequence, the study was conducted in communities that had more children eligible for Head Start than could be served with the existing number of funded slots.

At each of the selected Head Start centers, program staff provided information about the study to parents at the time enrollment applications were distributed. Parents were told that enrollment procedures would be different for the 2002-03 Head Start year and that some decisions regarding enrollment would be made using a lottery-like process. Local agency staff implemented their typical process of reviewing enrollment applications and screening children for admission to Head Start based on criteria approved by their respective Policy Councils. No changes were made to these locally established ranking criteria.

Information was collected on all children determined to be eligible for enrollment in fall 2002, and an average sample of 27 children per center was selected from this pool: 16 who were assigned to the Head Start group and 11 who were assigned to the non-Head Start group. Random assignment was done separately for two study samples—**newly entering** 3-year-olds (to be studied through two years of Head Start participation, kindergarten, and 1st grade) and **newly entering** 4-year-olds (studied through one year of Head Start participation, kindergarten, and 1st grade).

The total sample, spread over 23 different states, consists of 84 randomly selected Head Start grantees/delegate agencies, 383 randomly selected Head Start centers, and a total of 4,667 newly entering children, including 2,559 in the 3-year-old group and 2,108 in the 4-year-old group.[5] No statistically significant differences were found between the children randomly assigned to the Head Start and non-Head Start groups, providing one of several indications that the initial randomization was accomplished with high integrity, necessary for the validity of the impact estimates.

Data Collection

- Baseline data were collected in fall 2002 with annual spring follow-ups through 2006, the end of 1st grade for the youngest children.
- Comparable data are being collected for both Head Start and non-Head Start children, including interviews with parents, direct child assessments, surveys of Head Start and non-Head Start teachers, interviews with center directors and other care providers, direct observations of the quality of various care settings, and care provider ratings of children.

Data collection began in the fall of 2002 and will continue through the spring of 2006, following children from age of entry into Head Start through the end of 1st grade. Comparable data are being collected for both Head Start and non-Head Start children, including interviews with parents, direct child assessments, surveys of Head Start and non-Head Start teachers, interviews with center directors and other care providers, direct observations of the quality of various care settings, and care provider ratings of children.

[5] The sample of 3-year-olds is slightly larger than the sample of 4-year-olds to protect against the possibility of higher study attrition resulting from an additional year of longitudinal data collection for the younger children.

To date, response rates have been very good, with 83 percent of parents completing interviews in fall 2002 and spring 2003, and assessments being completed for 82 percent of the children. There is some difference in response rates between the Head Start and non-Head Start groups. Statistical weighting has been used both to adjust for the observed non-response and to generalize the data to the national Head Start program.

Statistical analysis of the characteristics of the sample used in this report (i.e., those children and parents for whom data were collected in spring 2003) indicate that the Head Start and non-Head Start groups are well matched on available characteristics, with only two small differences for each of the two age groups. These differences are not fully accounted for by the use of non-response adjustments to the sampling weights and are instead dealt with through their inclusion as covariates in the statistical models used to estimate program impacts.

Although every effort was made to ensure complete compliance with random assignment, some children accepted into Head Start did not participate in the program (this is not an uncommon occurrence in the program), and some children assigned to the non-Head Start group nevertheless entered the program, typically at centers that were not in the study sample. Statistical procedures for dealing with these events are discussed in the report. The findings in this report provide estimates of both the impact of access to Head Start using the sample of all randomly assigned children and a preliminary look at the impact of Head Start on program participants (adjusting for the deviations from random assignment).

Analysis Methods

Impact estimates discussed in this report represent the effect of Head Start on children and parents after one year of program participation.[6] Estimates are primarily based on the use of statistical models that control for any random differences in background characteristics between the Head Start and non-Head Start groups. Impacts are presented both for the overall average effects (for the full sample) and for selected subgroups of children and parents. All estimates use weighted data to generalize the findings to the full population of newly entering Head Start children.

[6] These are the average impacts of access to Head Start, often referred to as "intent to treat" impact estimates. Additional analysis on the children and parents who actually participated in the program (referred to as the "impact on the treated") are presented in appendices 4-8.

Before describing the results, three points are worth emphasizing.

1. **<u>The initial analyses represent only a portion of what is planned for future reports</u>:** In looking at child experiences, the current report provides only a partial set of preliminary indicators. Future reports will expand upon the description of the characteristics of the child care settings used by families and explore how child impacts vary with the quality of their early care experience. Additionally, future reports will address an expanded array of outcomes, the impacts of full-day/part-day programs, and other factors that have been shown to influence children's school readiness, such as teacher characteristics.

2. **<u>The non-Head Start (control) group is not a "no service" group</u>:** Parents of children in the control group were not precluded from enrolling their children in other types of preschool or child care arrangements. Consequently, the impact of Head Start is being evaluated against a mixture of alternatives available in the community, ranging from parent care to non-Head Start center-based programs. In some cases, these alternatives may look very much like Head Start, while others may look very different from Head Start. Evaluating Head Start against the current mixture of alternative arrangements isolates the contribution the Federal program is making relative to the array of other child care services currently available to low-income families.

3. **<u>The magnitude of estimated impacts must be viewed in context</u>:** This report uses a strict standard for reporting statistical significance. Only those impacts that could be detected with 95 percent confidence are reported as statistically significant. For those outcomes where statistically significant impacts were detected, results are provided in both their "natural" units (e.g., as points on a test score) and as "effect sizes" which provide a common yardstick for comparing across the different outcomes as well as to other research studies. When no significant impact was detected, effect sizes are not reported. For this report we have adopted the following conventions for interpreting effect sizes. Effect sizes of less than 0.2 are considered small, between 0.2 and 0.5 are considered a moderate impact, and over 0.5 are considered large impacts. For the most part, effect sizes from the current analysis are in the range of small to moderate. In considering the effect sizes, readers should keep in mind that:

   a. These findings represent the impact of Head Start after a single year of participation.

   b. There were some deviations from perfect random assignment that may affect the size and statistical significance of estimated impacts.

   c. Any judgment about the importance of the reported impact estimates must consider both the level of gains that children can be expected to achieve within a relatively short period of time and the size of effects that have been found in other early childhood and educational research studies.

Key Findings

As a way to provide a context for understanding the estimated program impacts, this section begins with a description of the early experiences of children assigned to the Head Start and non-Head Start groups. The impact findings are then organized by the two overarching research questions: (1) overall national average impacts on children's school readiness and parenting practices that support their development and (2) program impacts for particular subgroups of children and parents.

Within these two broad categories, results are organized by four outcome domains: (1) children's cognitive development, (2) children's social-emotional development, (3) children's health status and access to health care, and (4) parenting practices. Within each domain, results are presented separately for children in the 3- and 4-year-old groups.

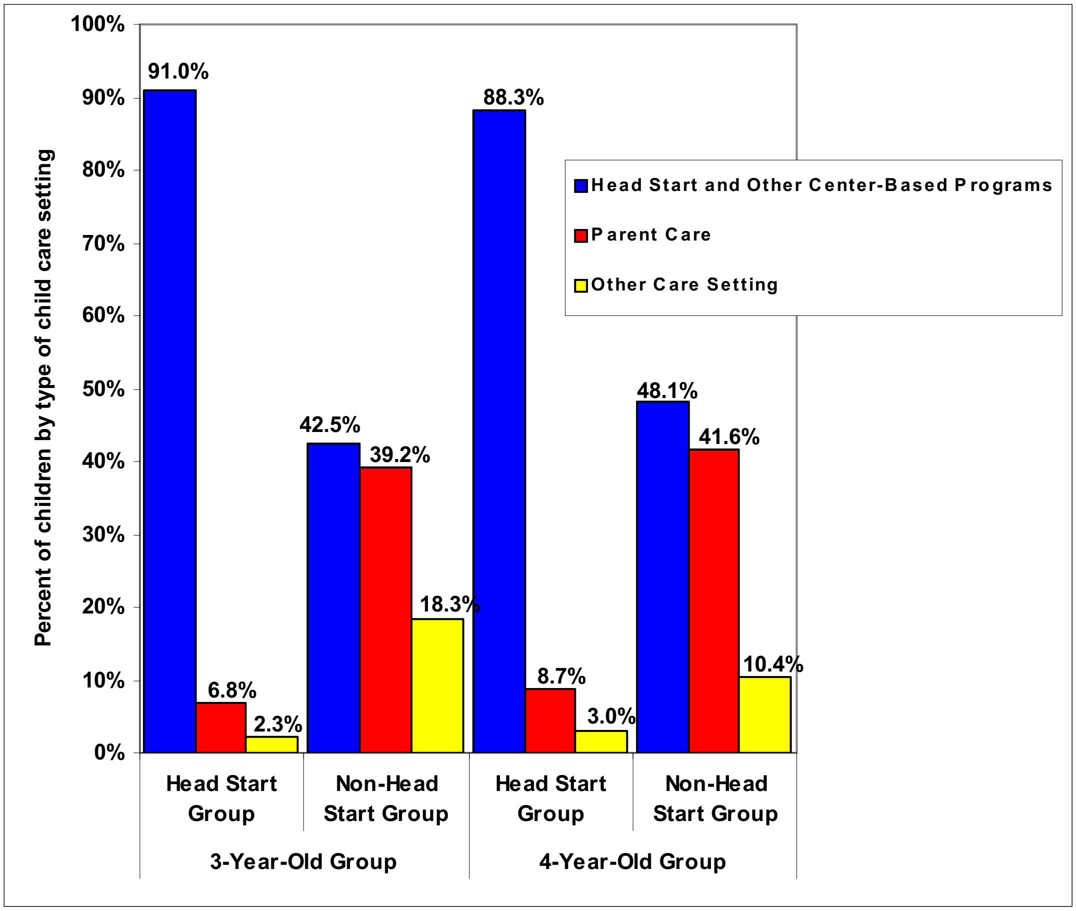
Children's Early Experiences

There is clear evidence that Head Start increases the likelihood that low-income children will be enrolled in center-based child care. Specifically, Head Start group children were twice as likely as the non-Head Start group children to use a center-based program in spring 2003. Approximately 90 percent of children in the Head Start group in both age cohorts were using a center-based program compared to 43 percent of children in the 3-year-old non-Head Start group and 48 percent of the 4-year-old non-Head Start group. Head Start group children were also more likely than non-Head Start group children to be in a center-based environment in **both** fall 2002 and spring 2003 and to have been in their spring 2003 setting since the start of the 2002-03 program year.

Conversely, non-Head Start group children were substantially more likely than Head Start group children to be exclusively in parent care[7] in spring 2003. Among children in the 3-year-old group, 39.2 percent of non-Head Start group children were in parent care as compared to only 6.8 percent of children in the Head Start group; among children in the 4-year-old group, the figures were 41.6 and 8.7 percent, respectively (see Exhibit 2).

[7] Exclusively in parent care is defined as being in **no** other non-parental setting for at least 5 hours per week.

Exhibit 2: Child Care Settings Used by Head Start and Non-Head Start Children, Spring 2003



The rates at which children in the study used Head Start or other center-based care did not differ substantially by age group. This is a somewhat surprising finding because in the general population, 4-year-olds are more likely than younger children to be enrolled in center-based programs.

In addition to conducting a preliminary examination of the impact of Head Start on children's use of early care arrangements, this report also presents findings on some initial quality indicators for the Head Start centers and other center-based programs attended by study children. These descriptive data provide some insight into the different environments in which Head Start and non-Head Start children are found when they attend centers, a difference that has important implications for understanding the impact of Head Start on children and parents. On the initial indicators assessed, children in the Head Start centers were in environments that more often (1) had positive interactions between children and teachers as measured by the Arnett Scale of

Teacher Behavior, (2) used curriculum and activities to enhance children's skills, and (3) had higher scores on the Early Childhood Environment Rating Scale: Revised Edition.

Overall Average Impacts

Impact on Children's Cognitive Development

Exhibit 3: Effect Sizes on Assessments for Which Head Start Had a Significant Overall Impact [1]

Cognitive Domains	Effect Sizes	
	3-Year-Old Group	4-Year-Old Group
Pre-Reading		
Woodcock-Johnson III Letter-Word Identification	**0.24**	**0.22**
Letter Naming	**0.19**	**0.24**
Pre-Writing		
McCarthy Draw-A-Design	**0.13**	--
Woodcock-Johnson III Spelling	--	**0.16**
Vocabulary		
PPVT-III Adapted	**0.12**	--
Color Naming	**0.10**	--
Parent Reported Literacy Skills	**0.34**	**0.29**
Oral Comprehension and Phonological Awareness	--	--
Early Math	--	--

[1] All effect sizes presented in table are based on statistically significant treatment and control differences of at least $p \leq 0.05$.

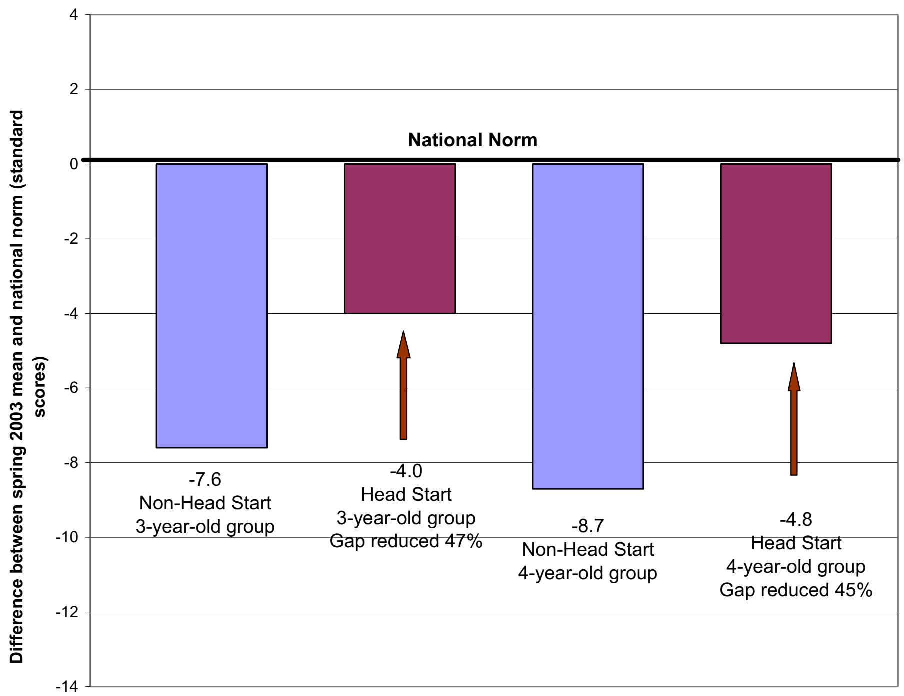
The impact of Head Start on children's cognitive development was examined in five constructs based on direct child assessments: (1) pre-reading skills focusing primarily on letter recognition, an important stepping stone on the path to becoming a proficient reader; (2) pre-writing skills that address children's ability at drawing shapes and writing letters; (3) vocabulary knowledge, which is indicative of children's receptive language development; (4) oral comprehension and phonological awareness which assess the ability to understand spoken language, including the knowledge that spoken sentences are made of component words that, in turn, comprise syllables and sounds (phonemes); and (5) early math skills that are essential for the development of more advanced quantitative capabilities. In addition, parents were asked to provide their perceptions of their child's emerging literacy and language skills.

As shown in Exhibit 3, the largest impacts were found for direct assessments of pre-reading skills and for parent-reported perceptions of their child's emergent literacy and language skills. Somewhat smaller impacts were found for the direct assessments of pre-writing skills and vocabulary (see Exhibit 3). No overall positive impact was found in the areas of oral comprehension and phonological awareness, or early math skills.

With regard to pre-reading skills, the effect sizes of the impacts on the Woodcock-Johnson III Letter-Word Identification test scores were 24 percent of a standard deviation for children in the 3-year-old group and 22 percent for children in the 4-year-old group. The effect sizes of the impact on the Letter Naming task were 19 percent for children in the 3-year-old group and 24 percent for children in the 4-year-old group.

Comparing the skill levels of children in the Head Start Impact Study with those of the general population of 3- and 4-year-olds in the United States (including those who were not from

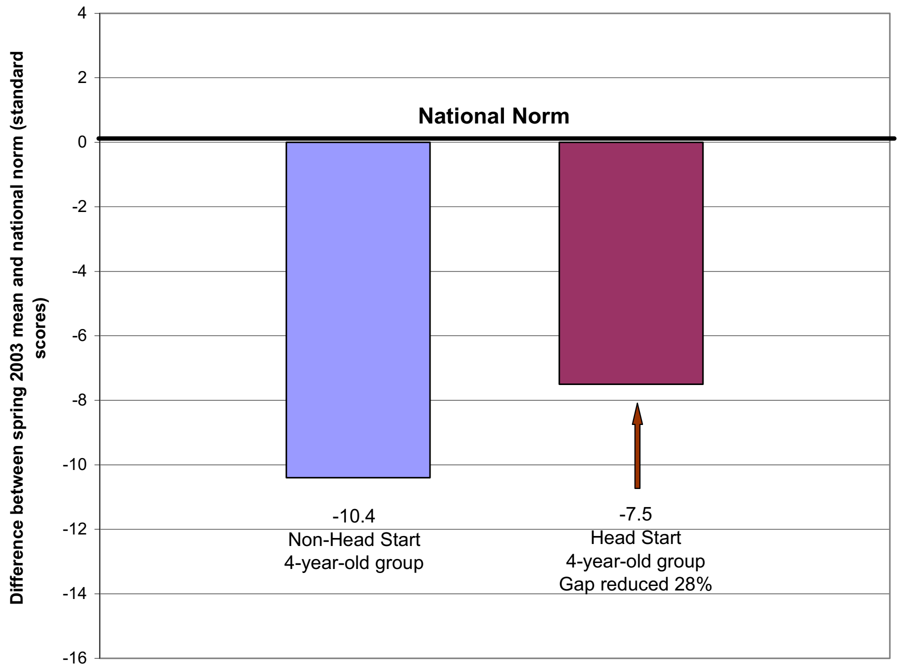
Exhibit 4: Impact of Head Start on Reducing the Achievement Gap in Children's Pre-Reading Skills (Woodcock-Johnson III Letter-Word Identification): Comparing Spring 2003 Means to National Norms by Age Group



low-income families) on the Woodcock-Johnson III Letter-Word Identification test showed that, after one year, the mean performance of Head Start children was still below the average performance level for all U.S. children, by about one-third of a standard deviation (about 5 points). However, at the end of one year, Head Start was able to nearly cut in half the achievement gap that would be expected in the absence of the program (as indicated by comparing the means for the Head Start and non-Head Start groups in Exhibit 4).

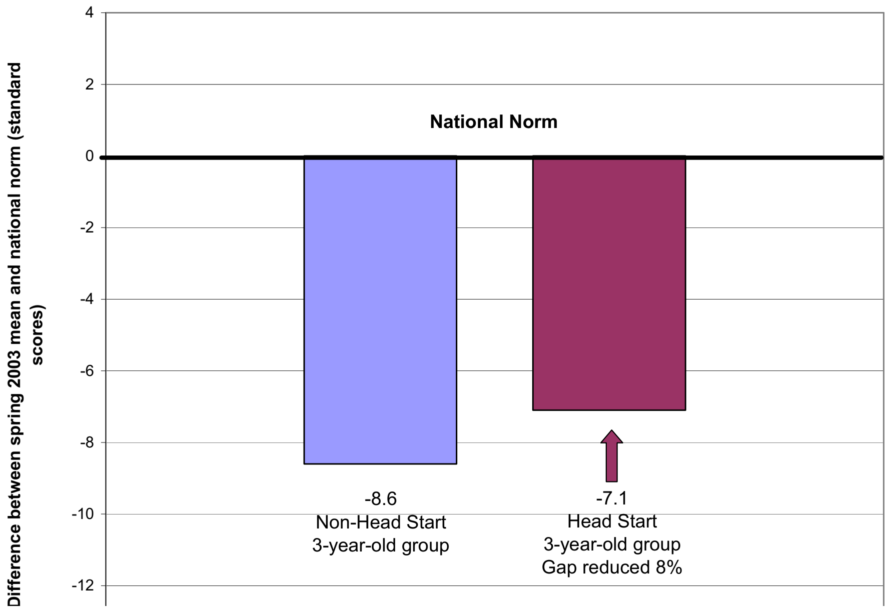
Among children in the 3-year-old group, the impact of Head Start on pre-writing skills was apparent in their score on the McCarthy Draw-a-Design test, which was 0.15 points higher for the Head Start group than the non-Head Start group with an effect size of 13 percent. For children in the 4-year-old group, there was also a positive impact on pre-writing skills for the Head Start group with an effect size of 16 percent as assessed by the Woodcock-Johnson III Spelling test. Head Start children were again found to be closer than non-Head Start children to the national norm for early writing skills by 28 percent (see Exhibit 5).

Exhibit 5: Impact of Head Start on Reducing the Achievement Gap in Children's Pre-Writing Skills (Woodcock-Johnson III Spelling): Comparing Spring 2003 Means to National Norms by Age Group



Statistically significant impacts on vocabulary knowledge were found, only for children in the 3-year-old group, with an effect size of 12 percent on the PPVT-III (Adapted) test. Thus, for this group only, Head Start children were 8 percent closer than non-Head Start children to the national norm on vocabulary skills (see Exhibit 6). No significant effects were found on vocabulary knowledge for the 4-year-old Head Start group.

Exhibit 6: Impact of Head Start on Reducing the Achievement Gap in Children's Vocabulary Skills (PPVT-III (adapted)): Comparing Spring 2003 Means to National Norms by Age Group



Impact on Children's Social-Emotional Development

The impact of Head Start on children's social-emotional development was examined along three dimensions: (1) social skills and positive approaches to learning that deal with curiosity, imagination, openness to new tasks and challenges, and having a positive attitude about gaining new knowledge and skills, (2) the incidence of various problem behaviors, and (3) social competencies.

Among children in the 3-year-old group, the frequency and severity of problem behavior reported by their parents were lower for children in the Head Start group compared to children in the non-Head Start group (see Exhibits 7 and 8). With regard to the overall problem behavior, the incidence of parent-reported problems was lower for 3-year-old children in the Head Start group (an effect size of 13 percent),

Exhibit 7: Effect Sizes for Social-Emotional Factors for Which Head Start Had a Significant Overall Impact [1]

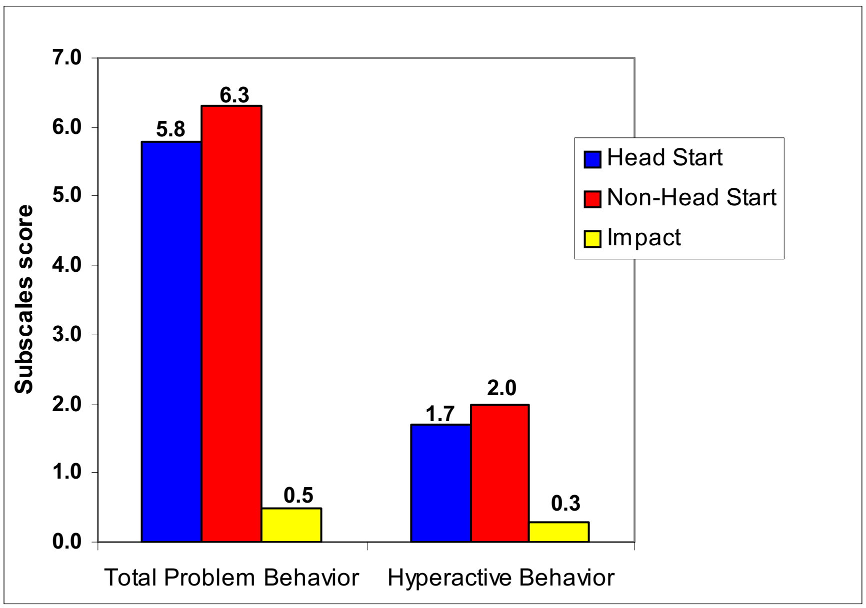
Social-Emotional	Effect Size	
	3-Year-Old Group	**4-Year-Old Group**
Problem Behaviors		
Total Behavior Problems	**-0.13**	**--**
Hyperactive Behavior	**-0.18**	**--**
Aggressive Behavior	--	--
Withdrawn Behavior	--	--
Social Skills and Approaches to Learning	**--**	**--**
Social Competencies	**--**	**--**

Negative effect sizes means reduction in problem behavior and aggressive behavior.
[1] All effect sizes presented in table are based on statistically significant treatment and control differences of at least p≤0.05.

and the incidence of parent report of hyperactive behavior was also lower for 3-year-old children in the Head Start group (an effect size of 18 percent). No overall impact of Head Start was found on the parent-reported Social Skills and Positive Approaches to Learning scale or on the parent-reported Social Competencies Checklist, for children in both age groups.

These measures are based on behavior reports from parents. An important additional source of information on children's social development—reports from children's teachers and caregivers—was not available for all children at this stage but will be available in future years of the study, when the children are in elementary school.

Exhibit 8: Impact of Head Start on Behavior Problems and Hyperactive Behavior, 3-Year-Old Group



Impact on Children's Health Outcomes

Head Start had a positive impact on certain indicators of children's health. The impact of access to Head Start on children's health was examined for a few selected measures reported by parents at the end of the first program year:

Exhibit 9: Effect Sizes for Health Care Factors for Which Head Start Had a Significant Overall Impact [1]

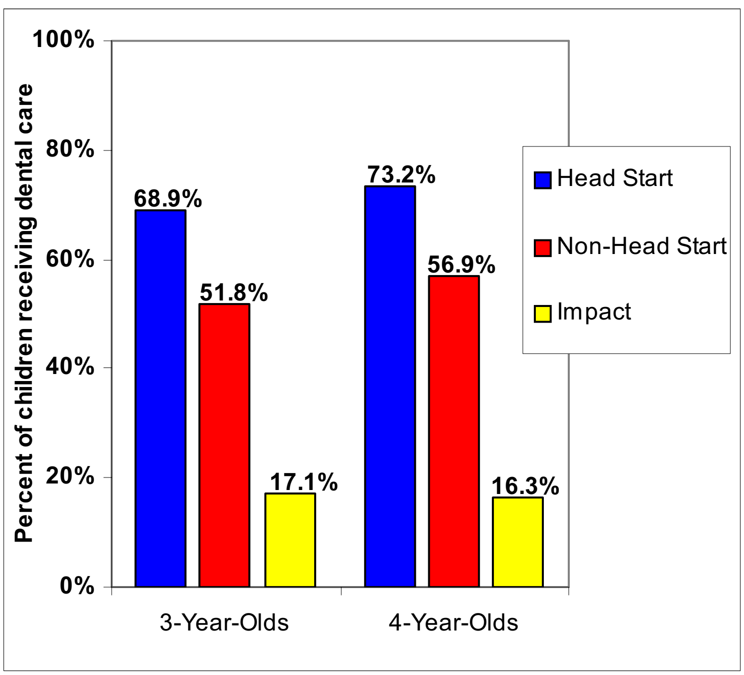
Health Outcomes	Effect Size	
	3-Year-Old Group	4-Year-Old Group
Access to Health Care		
Child Had Dental Care	**0.34**	**0.32**
Child Has Health Insurance	--	--
Health Status		
Overall Health Status	**0.12**	**--**
Child Needs Ongoing Care	**--**	**--**
Child Had Injury	**--**	**--**

[1] All effect sizes presented in table are based on statistically significant treatment and control differences of at least $p \leq 0.05$.

(1) the child's health status, including parent's report of the child's overall health status, whether the child needs ongoing care for an illness or condition, and whether the child had an injury in the last month and (2) the child's access to health care services, including whether the child has health insurance and whether the child has received dental care. No direct measures of children's actual health status, or their receipt of health care services, were undertaken for this study. Instead, data are based on parent report.

For children in both the 3- and 4-year-old group, a positive impact was found on the receipt of dental care (see Exhibits 9 and 10). The impact was similar for children in both age groups (17 percentage points for the 3-year-old group and 16 percentage points for the 4-year-old group), with similar effect sizes as well (34 percent and 32 percent, respectively). For children in the 3-year-old group, a positive impact was also found on parents' reported ratings of their children's health status, with more parents of children in the Head Start group reporting that their child's health was either excellent or very good (an effect size of 12 percent).

Exhibit 10: Impact of Head Start on Parent-Reported Receipt of Dental Care, 3- and 4-Year-Old Groups



Impact on Parenting Practices

One of the hallmarks of Head Start is its focus on parents as their child's first and primary teacher, recognizing that the involvement of parents is crucial for fostering children's school readiness. Historically, Head Start programs have reached out to families in a variety of ways, by encouraging parent involvement in their child's classroom, providing parent education to help strengthen parents' childrearing knowledge and skills, and providing referrals to address family needs so that parents can be more effective in their role as caregiver.

The impact of Head Start on parenting practices was examined in three main areas for this report: (1) educational activities that parents do with their children, including parent-child interactions that involve talking, reading, teaching, and exposure to new experiences that are crucial for promoting language development and early literacy; (2) parental discipline that emphasizes establishing firm but fair expectations for child behavior and promotes the development of social understanding and skills necessary for positive relationships with peers and adults; and (3) safety practices--parents' preventive efforts to safeguard the child's environment that are crucial for children's physical health and overall well-being.

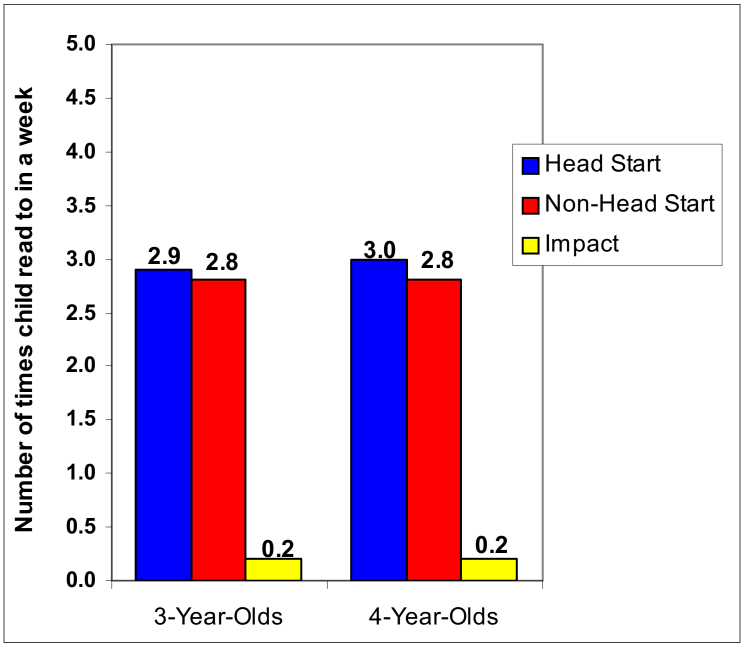
For both age cohorts, Head Start had a small positive impact on the extent to which parents reported reading to their child (see Exhibits 11 and 12), with an 18 percent effect size for the 3-year-old group and a 13 percent effect size for the 4-year-old group. Positive impacts also were found for children in the 3-year-old group on the extent to which their parents exposed them to a variety of cultural enrichment activities such as taking them to a museum or a zoo (an effect size of 11 percent).

Exhibit 11: Effect Sizes for Parenting Practices for Which Head Start Had a Significant Impact [1]

Parenting Practices	Effect Size	
	3-Year-Old Group	**4-Year-Old Group**
Educational		
Number of Times Child Read To	**0.18**	**0.13**
Family Cultural Enrichment Scale	**0.11**	**--**
Discipline Strategies		
Spank Child in Last Week	**-0.14**	**--**
Number of Times Spanked	**-0.10**	**--**
Use Timeout	**--**	**--**
Number of Timeouts	**--**	**--**
Child Safety Practices		
Overall Parental Safety Practices	**--**	**--**
Removing Harmful Objects	**--**	**--**
Restricting Child Movement	**--**	**--**
Safety Devices	**--**	**--**

Negative effect size reflects reduction in outcome.
[1] All effect sizes presented in table are based on statistically significant treatment and control differences of at least $p \leq 0.05$.

Exhibit 12: Impact of Head Start on the Number of Times Parent Reads to Child in a Week, 3- and 4-Year-Old Groups



For parents of children in the 3-year-old group, there is a lower use of physical discipline with children in the Head Start group compared to children in the non-Head Start group. A similar impact was not found on physical discipline for parents of children in the 4-year-old group. No statistically significant impacts were found on parents' child safety practices at home, for either age group.

Variation in Program Impact

It is important to understand how the impact of Head Start may vary among different types of children, parents, and communities and in relation to children's early childhood experiences. To fully understand these issues, it is necessary to assess both the **difference** in impact between subgroups (e.g., Does Head Start have larger effects on boys compared to girls?) and the impact of Head Start **on the individual subgroups themselves** (e.g., Does Head Start have an impact on boys?). To date, only an initial examination of sources of variation in program impacts has been undertaken; future reports will address this topic in more depth.

The analyses discussed in this report examine impacts on subgroups, and differences in impacts, for subgroups defined by the following child or parent characteristics: child gender, race and ethnicity; presence of special needs; and for only the cognitive outcomes, the child's status at the time of entry into Head Start; parent's marital status; age of mother at first birth; and primary caregiver's depressive symptoms. Positive impacts were found for a variety of subgroups of children with a range of demographic and family characteristics:

- **Child and home language:** For children in the 3-year-old group whose primary language was English, positive impacts were found on a variety of cognitive outcomes, as well as on particular measures of social-emotional development, health, and parenting practices. Among children in this age group whose primary language was Spanish, impacts were found across several domains but were fewer in number. For children in the 4-year-old group whose primary language was English, positive impacts were found in all domains; for children whose primary language was Spanish in this age group, impacts were found only in the area of health.
- **Race and ethnicity:** For children in the 3-year-old group, race and ethnicity appear to influence the extent of Head Start's impact, with particularly positive impacts noted in several domains for African American and Hispanic children. For the 4-year-old group, fewer impacts were found for minority children; observable impacts were particularly scarce for Hispanic children, a group found to have just one statistically significant impact (in the area of health).
- **Primary caregiver's depressive symptoms:** For children in the 3-year-old group, cognitive impacts were found to **decrease** with increasing levels of primary caregiver's reported baseline depressive symptoms. For children in the 4-year-old group, impacts were found to be sensitive to baseline depression for just one outcome, parent-reported child social competencies.
- **Age of mother at first birth:** In the 3-year-old sample, Head Start reduced the use of physical discipline when children misbehaved for mothers who had first given birth **before** the age of 19. In both the 3- and 4-year-old group, Head Start led mothers who had first given birth **after** the age of 19 to spend more time reading to their children, and to take them to a greater variety of cultural enrichment activities.

Contents of Report

This report, consisting of two volumes, presents early estimates of the impact of Head Start; however, much is yet to be done in this complex study to explore all the possible questions of policy and program interest.

Volume 1 consists of eight chapters. Chapter 1 presents the study background, including an overview of the study objectives, sample design, data collection, and response rates. Chapter 2 provides further details about the study sample, including a description of child and parent characteristics measured before and after random assignment. To provide a context in which to understand the impact findings, Chapter 3 examines the impact of Head Start on the types of preschool and child care settings that parents selected for their children as well as descriptive information on the characteristics of different types of early care arrangements. Chapter 4 presents an overview of the methods used for analyzing impacts on children and families.

The remaining four chapters present the results of the impact analyses. The impact of Head Start on children's cognitive development is presented in Chapter 5, focusing on six different domains of cognitive outcomes (i.e., pre-reading skills, pre-writing skills, vocabulary knowledge, oral comprehension and phonological awareness, early math skills, and parent report of children's literacy). The impact of Head Start on children's social-emotional development is presented in Chapter 6, focusing on parent-reported measures of social competencies, positive approaches to learning, and problem behaviors. Chapter 7 presents findings on the impact of Head Start on children's health status and access to health services, and Chapter 8 presents findings on the impact of Head Start on parenting practices in the areas of educational activities, discipline practices, and child safety practices. There are also technical appendices that present further details about the study design, the study sample, and analytic techniques.

Chapter 1: Study Background

The Congressional Mandate

Since its beginning in 1965, Head Start's goal has been to boost the school readiness of low-income children. The premise underlying the program is that low-income children do not receive the same level of intellectual stimulation at home as middle-class children. Based on a "whole child" model, the program provides comprehensive services that include preschool education; medical, dental, and mental health care; nutrition; and parental involvement. Head Start services are designed to be responsive to each child's and family's ethnic, cultural, and linguistic heritage.

In the late 1990s, the US General Accounting Office (GAO) released two reports concluding that (1) ". . .the Federal government's significant financial investment in the Head Start program, including plans to increase the number of children served and enhance the quality of the program, warrants definitive research studies, even though they may be costly"[1] and (2) this information need could not be met because ". . .the body of research on current Head Start is insufficient to draw conclusions about the impact of the national program."[2]

Based on these reports, and on the testimony of research methodologists and early childhood experts, Congress included in the 1998 reauthorization of Head Start a mandate that the US Department of Health and Human Services (DHHS) determine, on a national level, the impact of Head Start on the children it serves.[3] As noted by the Advisory Committee on Head Start Research, this legislative mandate requires that the impact study address two broad research questions:[4]

> "What difference does Head Start make to key outcomes of development and learning (and in particular, the multiple domains of school readiness) for low-income children? What difference does Head Start make to parental practices that contribute to children's school readiness?"
>
> "Under what circumstances does Head Start achieve the greatest impact? What works for which children? What Head Start services are most related to impact?"

---

[1] US General Accounting Office. (1998). ***Head Start: Challenges in Monitoring Program Quality and Demonstrating Results.*** Washington, DC: Author.

[2] US General Accounting Office. (1997). ***Head Start: Research Provides Little Information on Impact of Current Program.*** Washington, DC: Author.

[3] See Appendix 1.1 for the research-related amendments to the *Head Start Act* included in the 1998 reauthorization.

[4] Advisory Committee on Head Start Research and Evaluation (1999). ***Evaluating Head Start: A Recommended Framework for Studying the Impact of the Head Start Program.*** Washington, DC: US Department of Health and Human Services.

The Advisory Committee set forth a framework for research on the impact of Head Start that is both scientifically credible and feasible. The Committee acknowledged that the legislative language recommended the use of a rigorous methodology, including random assignment of children to Head Start and non-Head Start groups at a diverse group of sites, selected nationally and reflecting the range of Head Start quality across the country. To implement this design, DHHS awarded a contract in October 2000 to Westat of Rockville, MD, in collaboration with The Urban Institute, the American Institutes for Research, and Decision Information Resources.

Study Objectives and Research Questions

The first broad research question noted above can be divided into two parts: (1) the **direct** effect of Head Start on children's early development and (2) the extent to which Head Start has an **indirect** effect by improving the ability of parents to support their children's learning and development. Though not specifically identified, it is also valuable to understand the extent to which Head Start may affect the nature, duration, and quality of children's early care and program experiences, which may, in turn, lead to improvements in school readiness. The second broad research question recognizes the importance of also understanding how the impact of Head Start may vary for different types of children and their families and in relation to the nature, duration, or quality of a child's early care and program experiences. These broad research questions led to the specification of the following more detailed study questions that have guided the design and implementation of the Head Start Impact Study:[5]

- **Impacts:**

  1. What impact does Head Start have on school readiness including children's approaches to learning, language development and emergent literacy, mathematical ability, physical well-being and motor development, and social and emotional development?

  2. What impact does Head Start have on parental practices that contribute to children's school readiness (e.g., time spent reading to their child)? To what extent are these parenting practices related to child development outcomes?

  3. What impact does Head Start have on the nature and quality of children's early care and program experiences (e.g., the intensity of reading instruction)? To what extent are these experiences related to child development outcomes?

---

[5] For more details on the design of the Head Start Impact Study see: Administration for Children and Families. (2003). ***The Head Start Impact Study: Research Design and Preliminary Analysis Plan.*** Washington, DC: US Department of Health and Human Services.

- **Variation in impacts for certain subgroups of children and families:**

  1. Do impacts vary according to children's characteristics at the time of entry into Head Start? Are there some subgroups that benefit while others do not? Subgroup characteristics include gender, race/ethnicity, age at program entry (3- vs. 4-year-olds), presence of disabilities, as well as the child's status on a number of developmental characteristics (e.g., language ability) at the point of Head Start entry.

  2. Do impacts vary by characteristics of the child's home environment at the time of entry into Head Start? What particular environments lead to positive impacts? Home characteristics include family structure (e.g., single parent, teen mother), household income, and parental practices related to school readiness before exposure to Head Start.

  3. Do impacts vary by the characteristics of the community where participants reside? In which types of communities does Head Start produce clear gains? Community characteristics include characteristics of the economic and social environment (e.g., poverty, unemployment rates), and the policy environment related to the availability and quality of alternative services for low-income children (e.g., state and local government funding for preschool programs).

- **Variation in impacts related to characteristics that may be affected by Head Start participation:**

  1. Do impacts vary by parent's ability to support their children's development or by characteristics of the home environment (e.g., does the frequency with which an adult reads to a child influence literacy outcomes)? Which subgroups based on at-home supports gain from Head Start participation?

  2. Do impacts on children vary by the nature, duration, and quality of their early care and program experiences? For example, do impacts vary by the amount of language instruction they receive?

Overview of the Study Design and Implementation

As discussed above, the primary purpose of this study is to determine whether Head Start has an impact on participating children and their parents and, if so, whether such effects vary among different types of children, families, communities, and configurations of children's early care and program experiences. By impact we mean a difference between the outcomes observed for Head Start participants and what *would have been observed for these same individuals had they **not** participated in Head Start*. This focus on impacts distinguishes this study from many others that seek primarily to examine relationships among participant outcomes and between participant outcomes and one or more individual or program characteristics (see, for example, the

Head Start Family and Child Experiences Study (FACES)[6]). Instead, the present study uses information from participants *and a statistically equivalent group of children who do not participate in Head Start* to determine whether Head Start **caused** the observed child and parent outcomes.

Given this goal of measuring program impacts, how do we determine what outcomes would have been observed if the children had not participated in Head Start? That is, how do we observe children having the same characteristics in two places at the same time—in Head Start and not in Head Start—and compare them? In many studies, researchers have addressed this problem by comparing program participants to a "participant-like" group of children who, in the ordinary course of events, do not participate in Head Start. However, even the best attempts at constructing such a comparable group of non-participants suffer from what evaluators call "selection bias." That is, families who seek out, or "select," Head Start for their children are likely to be different from those who do not on important factors that may lead to different outcomes independently of the effect of Head Start services. For example, parents who apply to Head Start may be more motivated to see that their children are well prepared to start school than those parents who choose not to seek Head Start enrollment. Moreover, the reasons why these parents make different decisions are both typically unobserved and likely to be related to the outcomes of interest in their own right. That is, the motivated parents do a host of things that may affect their children's development beyond enrolling them in Head Start. Because all of these differences cannot be accounted for, there is a risk of misattributing to program participation observed differences on a particular outcome measure (e.g., emergent literacy) that may be a result of intrinsic differences between participant and non-participant families.

To avoid this problem of selection bias, the Head Start Impact Study **randomly assigned** a sample of 3- and 4-year-old Head Start applicants not previously served by the program,[7] either to a **treatment group** (in which children and families received Head Start services) or to a **control group** (in which children were not granted access to Head Start but may have received a range of other services chosen by their parents). Under this randomized design, a simple comparison of outcomes for the two groups yields an unbiased estimate of the impact of Head Start on children's school readiness. The advantage of this research design is that if random

[6] US Department of Health and Human Services, Administration for Children, Youth and Families. (2003). ***Head Start FACES 2000: A Whole-Child Perspective on Program Performance, Fourth Progress Report.*** Washington, DC: Author.

[7] The Head Start Impact Study focuses on newly entering children to ensure that the estimated impacts are unaffected by previous program participation. Consequently, children who were returning to Head Start, as well as those previously enrolled in Early Head Start, were excluded from the study sample.

assignment is properly implemented with a sufficient sample size, program participants should not differ in any systematic or unmeasured way from non-participants except through their access to Head Start services.[8]

Sample Selection

Most randomized studies are conducted in small demonstration programs or, if done in an ongoing program, only in a small number of operating sites, usually those that volunteer to be included in the research. In contrast, the Head Start Impact Study is based on a nationally representative sample of both Head Start programs and newly entering 3- and 4-year-old children. That is, children applying for entry into Head Start in fall 2002, from a nationally representative sample of programs, were selected at random. This makes results generalizable to the entire Head Start program, not just the selected study sample. This approach responds to the congressional mandate and recommendations of the Advisory Committee that the study provide *"a national analysis of the impact of Head Start"* based on the selection of Head Start grantee/delegate agencies[9] that *"operate in the 50 states, the Commonwealth of Puerto Rico, or the District of Columbia and that do not specifically target special populations"* and that also reflect variation on a variety of characteristics, including *"region of the country, race/ethnicity/language status, urban/rural, and depth of poverty in communities,"* and *". . .design of program as a one-year or two-year experience for children; program options (e.g., center-based, home-based, part-day, full-day); auspice (e.g., Community Action Agency, public school, non-profit organization); community-level resources; alternative child care options for low-income children; and, the nature of the child care market and the labor market in the community studied."*

To meet these requirements, the study used a multi-stage sampling process to select a representative group of Head Start programs. The process, depicted in Exhibit 1.1, is described below:

1. **Initial grantee/delegate agency selection.** The sampling process began by using the Head Start Program Information Report (PIR) to create a list of 1,715 Head Start grantee and delegate agencies operating in fiscal year (FY) 1998-1999, after excluding (1) grantee/delegate agencies serving **only** special populations (migrant and tribal Head Start programs and sites serving **only** Early Head Start children), (2) grantees involved in the FACES 2000 study, and (3) as

[8] More precisely, there will be differences between individuals in the two groups, but the expected or average value of these differences is zero except through the influence of Head Start (i.e., selection bias is removed by random assignment).

[9] The study sample includes both Head Start grantees and their delegate agencies. Grantees are organizations that have fiscal and administrative responsibility for programs in their jurisdiction. In some cases, they can subcontract with agencies to handle administrative oversight over some or all of these programs. Throughout this report we use the term grantee/delegate agencies to refer to both types of agencies.

Exhibit 1.1: Sample Selection Process for the Head Start Impact Study

All FY1998-1999 Head Start Grantee/Delegate Agencies in All 50 States, DC, & Puerto Rico
Exclude "very new," migrant, Tribal Organization, and Early Head Start-only grantee/delegate agencies (N=1,715).

↓

Create Geographic Grantee Clusters and Group into 25 Strata
Group grantee/delegate agencies by geographic proximity with a minimum of 8 per cluster (N=161 clusters). Stratify clusters on: state pre-K and childcare policy; child race/ethnicity, urban/rural location, and region. Select 1 cluster per stratum with probability proportional to Head Start enrollment (N=261 grantee/delegate agencies).

↓

Determine Eligible Grantee/Delegate Agencies in Each Cluster
Exclude closed or merged programs and those that are "saturated" (i.e., have very few unserved children in the community). Eliminated 38 grantee/delegate agencies (N=223). Small grantee/delegate agencies were then grouped to ensure meeting target sample sizes (N=184 groups).

↓

Stratify and Select Grantee/Delegate Agencies
Stratify on grantee/delegate agency characteristics and local contextual variables, and randomly select approximately 3 grantee/delegate agencies per cluster (N=76 grantee groups, 90 grantee/delegate agencies across 23 states).

↓

Recruit Grantee/Delegate Agencies for the Study
Resulted in 76 grantee/delegate agency groups and 87 individual grantee/delegate agencies.

↓

Develop List of Head Start Centers
Participating grantee/delegate agencies provided lists of operating centers as of fall 2002 (N=1,427 centers).

↓

Determine Eligible Centers and Create Center Groups
Exclude saturated centers and create center groups by combining small centers with nearby centers (N=1,258 centers).

↓

Stratify and Select Sample of Centers
Stratify centers using same characteristics used with grantees. Randomly select centers and exclude saturated centers (84 grantee/delegate agencies, 383 centers).

↓

Select Children and Conduct Random Assignment
Final Sample: 84 grantee/delegate agencies, 378 centers, 2,783 Head Start children and 1,884 non-Head Start children.

recommended in the Advisory Committee report (1999), grantees/delegate agencies that were "*extremely new to the program*."[10] This pool of 1,715 Head Start programs was subsequently organized into 161 "geographic clusters" (to increase the ability to closely monitor random assignment and obtain high quality data) and then grouped into 25 strata to control for factors such as region of the country, urban/rural location, race/ethnicity, and variation in state pre-kindergarten and child care policies. One cluster of programs was then randomly selected from each of the 25 strata with probability proportional to total enrollment, providing a total of 261 grantee or delegate agencies in the sampled clusters (to improve efficiency, random subsampling was done in three very large urban clusters).

2. **Determining grantee/delegate agency eligibility.** To be eligible for inclusion in the study sample, grantee/delegate agencies had to have enough "extra" or additional newly entering applicants beyond their number of funded slots to allow for the creation of a non-Head Start control group. That is, the programs could not be serving all the eligible children in their community who wanted Head Start, a situation we refer to as "saturation." Ethically, random assignment could only be conducted in communities where Head Start programs were expected to be unable to serve all the eligible children seeking enrollment for fall 2002. This eligibility was determined from information verified through telephone calls to all 261 grantee/delegate agencies, augmented with information provided by Federal Regional Office staff and with data obtained from secondary sources such as local Child Care Resource and Referral Agencies, and the PIR. This screening process eliminated 28 grantees/delegate agencies (a reduction of 11 percent) found to be operating in saturated communities. Additionally, 10 other grantee/delegate agencies had been closed or merged, further reducing the pool of eligible programs to 223 grantee/delegate agencies.

3. **Selecting grantee/delegate agencies.** To ensure the inclusion of the full range of Head Start grantee/delegate agencies, smaller programs were combined with other agencies in the same cluster to form "grantee/delegate agency groups." These groups (some of which consisted of a single grantee or delegate agency) were then stratified along several dimensions: urban location (central city, other urban, rural/small town), auspice (school based versus all other agency types), percentage Hispanic and percentage African American enrollment, program options offered (part-day only, full-day only, both), and the percentage of total enrollment represented by newly entering 3-year-olds. Approximately three grantee/delegate agency groups were randomly selected from each of the 25 strata with probabilities proportional to the number of newly entering children. This yielded a sample of 76 grantee/delegate agency groups comprising 90 individual grantee/delegate agencies, across 23 states.

4. **Grantee/delegate agency recruitment.** Senior project staff visited all 90 selected grantee/delegate agencies during summer 2001 to explain the study, verify information needed for study implementation, and to gain their agreement to participate in the Head Start Impact Study. Three agencies were dropped at this point—one had recently closed and two were dropped due to an overlap with a study being conducted by the federally funded Head Start Quality Research

[10] Defined as in operation for fewer than 2 years.

Center—leaving 87 grantee/delegate agencies in 76 grantee/delegate agency groups (i.e., the overall number of grantee/delegate agency groups was not reduced).

5. **Identifying operating Head Start centers.** Because administrative data do not identify individual Head Start centers, each of the 87 grantee/delegate agencies was asked to provide a list of all centers expected to be in operation for the 2002-03 program year and to validate basic data about the characteristics of children served, program options, and enrollment patterns in each center. This resulted in a list of 1,427 Head Start centers in the 87 grantee/delegate agencies (76 grantee groups) that could potentially be included in the Head Start Impact Study.

6. **Determining center eligibility and selecting a sample of study centers.** The center-level data were first used to eliminate 169 centers determined to be "saturated," as was done previously for grantee/delegate agencies. This step reduced the total eligible pool of centers from 1,427 to 1,258 across 84 separate grantee/delegate agencies in 76 grantee/delegate agency groups (a reduction of about 11 percent and the loss of three grantee/delegate agencies, but no grantee groups). Next, small centers were combined with nearby centers, and the resulting "center groups" were then stratified using the same characteristics used for the selection of grantee/delegate agencies (excepting those that do not vary within grantee/delegate agencies such as a region). A main sample consisting of an average of three center groups was selected from each eligible grantee/delegate agency, resulting in a main sample of 448 centers in 84 grantee/delegate agencies.

   More in-depth or up-to-date information on the initially sampled centers led to a determination that some were, in fact, ineligible for inclusion in the study. These included centers that: (1) had recently closed or had been merged with other centers; (2) served only Early Head Start children; (3) were collaborations between Head Start and private preschool programs that could not subject their entire pool of applicants to random assignment; or (4) were, in fact, saturating their community with Head Start services. These findings resulted in the dropping of 103 initially sampled centers, but the addition of 38 replacement centers[11] to yield a sample of 383 centers.

This sample of Head Start grantee/delegate agencies and centers, when properly weighted (see Appendix 1.2), was designed to yield a sample of children that represents the national population of newly entering children and their families (with the exclusions noted above).

Random Assignment

At each of the selected Head Start centers, program staff provided information about the study to parents at the time enrollment applications were distributed. Parents were told that

[11] A "reserve" sample of an average of two center groups per program (a total of 237 centers) was also selected to be used as replacement sites if needed to achieve the expected overall study sample size of children. Thirty-eight of these centers were used. The final sample was 383 (448-103+38) centers.

enrollment procedures would be different for the 2002-03 Head Start year and that some decisions regarding enrollment would be made using a “lottery-like” process.[12]

The study team assigned local site coordinators to work with grantee/delegate agencies in each of the 25 geographic clusters to ensure that parents received this information with their applications. These site coordinators were also responsible for obtaining data on **all** applications for the 2002-03 program year (to ensure equal treatment of all applicants) and listing these data on a roster that was subsequently key-entered by central office study staff. Returning children, and a small number of grantee-requested “high-risk” exclusions,[13] were eliminated from these lists, and checks were made for duplicate records. The high-risk exclusions were made on a case-by-case basis with each grantee/delegate agency and in close consultation with Administration for Children and Families staff. Examples of such exclusions included children of homeless families, children in families with documented abuse and neglect, and children with severe disabilities, especially those disabilities that would make it difficult to test these children and include them in the study sample (e.g., blindness). Each grantee was limited to one exclusion per center for its total pool of newly entering children. In fact only 276 exclusions were taken out of a total of approximately 18,000 newly entering applications.

At this point, local agency staff implemented their typical process of reviewing enrollment applications and screening children for admission to Head Start based on criteria approved by their respective Policy Councils. No changes were made to these locally established admission criteria. Site coordinators recorded basic information about each applicant and what was usually a numerical score determined by local staff that signified the relative need of individual children (e.g., in some agencies, a higher score indicated a greater need for Head Start and a corresponding higher priority for admission). Using these rankings, the list of newly entering children who would ordinarily have been enrolled was “extended” to add a specified number of children needed for the non-Head Start control group. The children added were those who would normally be “next in line” for admission if the initially targeted children could not be enrolled.

[12] Children randomly assigned to the non-Head Start group were not to be admitted to Head Start during 2002-03. Those who were in the 3-year-old group, however, were told that they could re-apply for Head Start in 2003-04 and may be admitted if eligible.

[13] This decision was made because: (1) there were ethical concerns about assigning very high-risk children to the non-Head Start group, especially in situations where Head Start may provide their only option for early childhood services; (2) a previously conducted study demonstrated that the potential exclusion of those most severely in need affected cooperation when trying to recruit study sites; and (3) there were some children who could not be assigned to the non-Head Start group because of placement by the local child welfare agency.

The goal was to randomly select, on average, 27 children from the expanded list at each of the sampled centers or center groups: 16 to be assigned to the Head Start group and 11 to be assigned to the non-Head Start group. For an average center group, the 11 non-Head Start control group children represented about 9 percent of total enrollment. Where necessary, stratification was used, such as in situations where the degree of saturation varied by program option (part-day versus full-day) or age cohort. In some cases, where fewer children than expected were actually available, a smaller sample of children was selected for the study.

The original legislative mandate required that the Head Start Impact Study "*to the extent practicable*" address possible variation in program impact related to "*the length of time a child attends a Head Start program (and) the age of the child on entering the Head Start program*." This requirement reflects the hypothesis that different program impacts may be associated with 1 versus 2 years of Head Start experience. It also reflects a trend of increased enrollment of 3-year-olds in some grantee/delegate agencies presumably due to the growing availability of preschool options for 4-year-olds (often state-sponsored programs). Consequently, the study included two separate samples: a **newly entering** 3-year-old group (to be studied through 2 years of Head Start participation, kindergarten, and 1st grade), and a **newly entering** 4-year-old group (to be studied through 1 year of Head Start participation, kindergarten, and 1st grade). The 3-year-old group is slightly larger than the 4-year-old group to protect against the possibility of higher study attrition resulting from an additional year of longitudinal data collection for the younger children.[14]

Within the final set of 76 grantee/delegate agency groups (or 84 total grantees/delegate agencies), random assignment was attempted at a total of **383 randomly selected Head Start centers**. Of these, random assignment could not be completed in only five centers (or 1.3 percent), resulting in a final sample of 378 centers with successful random assignment. However, as noted above, the full desired sample could not be obtained at each center, resulting in the following situations:

- **Obtained Full Sample.** Random assignment was completed at 173 Head Start centers that provided the full expected sample of children.

[14] This roughly equal sampling of 3- and 4-year-old applicants was done despite the fact that 4-year-olds represent about twice the proportion of all Head Start participants as do 3-year-olds. In large part, this is because the 4-year-olds include both newly entering 4-year-olds plus returning children who began Head Start as 3-year-olds and who have turned 4 years of age in their second year of program participation.

- **Obtained Smaller Sample.** Random assignment was completed at 150 Head Start centers that provided a smaller than expected sample (i.e., because new application rates were lower than estimated).

- **Obtained Larger Sample.** Random assignment was completed at 55 Head Start centers that provided a larger than expected sample (i.e., because application rates for newly entering children were higher than originally estimated, sample sizes were increased to compensate for other centers that were unexpectedly low).

In total, 4,667 newly entering children were randomly assigned and included in the Head Start Impact Study (see Exhibit 1.2).

Exhibit 1.2: Number of Children in the Head Start and Non-Head Start Groups, by Age Group

Age Group	Head Start (Treatment) Group	Non-Head Start (Control) Group	Total Sample
3-year-olds	1,530	1,029	2,559
4-year-olds	1,253	855	2,108
Total	**2,783**	**1,884**	**4,667**

As indicated above, about 60 percent of the sample was assigned to the Head Start group and about 40 percent was assigned to the non-Head Start group. This imbalance reduces the precision of the impact estimates by less than 2 percent (compared to a balanced 50-50 design). However, it provided several important benefits: (1) it significantly increased the ability to recruit Head Start grantees and centers by decreasing the number of extra children needed for the control group, (2) decreased the loss of sites due to saturation, and (3) saved considerably on data collection costs because treatment group members (who participate in Head Start) require less effort to track and interview over time than children in the non-Head-Start control group.

Data Collection

Data collection began in fall of 2002 and will continue through the spring of 2006, following children from age of entry into Head Start through the end of the preschool years, end of kindergarten, and end of 1st grade. Comparable data are being collected for both Head Start and non-Head Start children and consist of the following:

- Measures of children's development that include (1) direct child assessments, (2) parent reports, and (3) teacher/care provider reports. Child outcomes are measured in the key domains of cognitive development (including assessment of skills in the areas of reading, writing, vocabulary, oral comprehension and phonological awareness, and math), social-emotional development, and health.

- Characteristics and quality of children's home environments are measured through (1) parental reports of beliefs and attitudes about their child's learning and parental participation in, and satisfaction with, their child's child care experience; (2) family household and demographic information, including parent-child relationships and the quality of the child's home life; (3) parent ratings of their child's behavior problems, social skills, and competencies; (4) parent's perceptions of their child's accomplishments; (5) parent's perception of their relationship with their child; and (6) child and family receipt of a variety of comprehensive services.
- Characteristics and quality of the primary preschool and child care arrangement as measured through (1) interviews with center-based directors, (2) surveys of teachers or interviews with care providers, and (3) observations of these settings.

To complete these data collection activities in the 25 geographic areas where children and families were sampled, the study uses measurement teams comprising local field interviewers/assessors and observers who work under the supervision of a site coordinator tasked with ensuring that all aspects of local data collection are completed during the field period.

This report focuses on the findings of the first year of data collection after random assignment of both Head Start and non-Head Start children. The following describes the data sources and measures used during the first year. Response rates and subsequent data collection plans are also summarized.

Fall 2002 Data Collection

Baseline data were collected in fall 2002 and included in-person interviews with the parent/primary caregiver of each study child and one-on-one child assessments conducted by the local interviewers/assessors:

- **Parent/Primary Caregiver Interviews.** Parent interviews were typically conducted in the child's home with a parent or primary caregiver living with and responsible for raising the child. Parent interviews were available in both English and Spanish versions, and bilingual English/Spanish speakers were hired for areas with Spanish-speaking families. For other languages, either interviewers/assessors fluent in these languages were hired or other local resources were asked to identify interpreters to aid in completing the parent interviews.

- **Child Assessments.** Child assessments provide direct measures of how well Head Start and non-Head Start preschool programs, or other child care, are achieving the goal of assisting children to be physically, socially, and educationally ready for success in kindergarten. The assessment battery (see Exhibit 1.3) is composed of a short series of tasks that are feasible and interesting for preschoolers to carry out and that have been shown to be predictive of later school success (test citations are provided in Appendix 1.4). The 35 to 45 minute child assessment battery was typically administered one-on-one by specially trained assessors in the child's "main" care setting, i.e., where the child spends the most time Monday through Friday between the hours of 9 AM and 3 PM.

  At the time of the assessment, the interviewer/assessor asked the main care provider a series of questions to determine the appropriate language for the assessment (see Appendix 1.3). For children requiring assessment in Spanish, a bilingual interviewer/assessor administered the assessment battery in Spanish and also administered two subtests in English, i.e., the Peabody Picture Vocabulary Test (adapted) (PPVT) and the Woodcock-Johnson III Letter-Word Identification. In spring 2003, the children assessed in Spanish in fall 2002 were assessed primarily in English, along with the continued administration of two Spanish language measures: the Test de Vocabulario en Imágenes Peabody (TVIP) and the Batería Woodcock-Muñoz Identificación de Letras y Palabras. One exception is Puerto Rico where, because instruction is in Spanish, all children were assessed only with the complete Spanish battery in spring 2003. For children who could not be assessed in either English or Spanish, a bilingual interviewer/assessor or an interpreter for the child's language were used. The interviewer/assessor (or interpreter) used the English assessment booklet, translated the instructions into the child's language, and administered four subtests: McCarthy Draw-A-Design, Color Names and Counting, Leiter-R-Adapted, and Story and Print Concepts. For the spring assessments, these children were all tested in English.

In addition, site coordinators visited all study Head Start centers to ascertain whether children assigned to Head Start were, in fact, attending and whether any control group children had been inadvertently enrolled in Head Start.

Fall 2002 data collection was completed by mid-November for the majority of children and parents (although a small number did extend into December). The implication of this late baseline data collection is discussed in Chapter 4, along with procedures used to deal with it in the analysis of program impact.

Winter 2003 Parent Updates

In the winter of 2003, a 10-minute telephone interview was conducted with parent/primary caregivers; in some instances, in-person interviews were conducted. These short

Exhibit 1.3: Direct Child Assessments – Fall 2002

Assessment Areas	Measure	English-Speaking Children	Spanish-Speaking Children*	Children Who Spoke Languages Other than English or Spanish
Language Development & Literacy	Peabody Picture Vocabulary Test III (PPVT III, adapted)	X	X	--
	Test de Vocabulario en Imágenes Peabody (TVIP, adapted)	--	X	--
	Comprehensive Test of Phonological and Print Processing (CTOPPP):			
	a. Print Awareness	X	--	--
	b. Elision	X	--	--
	Comprehensive Test of Phonological and Print Processing (CTOPP)(Spanish version):			
	a. Print Awareness	--	X	--
	b. Elision	--	X	--
	Woodcock-Johnson III:			
	a. Oral Comprehension	X	--	--
	b. Letter-Word Identification	X	X	--
	c. Spelling	X	--	--
	Woodcock-Muñoz R:			
	a. Identificación de Letras y Palabras	--	X	--
	b. Dictado	--	X	--
	Story and Print Concepts ***	X	X	X
	Letter Naming Task **	X	X	X
Mathematics	Counting Bears Task ***	X	X	X
	Woodcock-Johnson III: Applied Problems	X	--	--
	Woodcock-Muñoz R: Problemas Aplicadas	--	X	--
Early Writing	McCarthy Draw-a-Design	X	X	X
	Woodcock-Johnson III: Spelling	X	--	--
	Woodcock-Muñoz R: Dictado	--	X	--
Other Cognitive Ability	Color Names ***	X	X	X
Sustained Attention	Leiter-R Attention Sustained Task (adapted) ***	X	X	X
Assessor Ratings	Task persistence	X	X	X
	Attention span	X	X	X
	Body movement	X	X	X
	Attention to directions	X	X	X
	Comprehension of directions	X	X	X
	Verbalization	X	X	X
	Ease of elationship	X	X	X
	Confidence	X	X	X

* The Spanish version uses the Woodcock-Muñoz R. In this version, the Dictation subtest is used in place of the Spelling subtest. Children in Puerto Rico were administered only the Spanish subtests.
** Administered only in spring 2003.
*** Subtest administered in the language in which the child was assessed.

interviews were designed to obtain up-to-date contact information and child care setting information critical to determining the appropriate setting for spring 2003 data collection.

Spring 2003 Data Collection

In spring 2003, the interviewers/assessors again conducted in-person parent interviews and child assessments. Additional information was obtained from in-person interviews with directors of the Head Start and non-Head Start centers that study children attended, and teachers and other care providers completed self-administered questionnaires to rate each of the study children who were in their classroom or care. Teachers also completed questionnaires, and care providers were interviewed in person, to obtain information about themselves, the nature of the setting in which they work, and the types of services they provide to the selected study children. To further measure quality of care, direct observations of classrooms and family day care homes were conducted. To obtain comparable information on quality across all care settings, a five-question observational instrument was completed in every care setting visited, including the child's own home. Each of these activities is described below:

- **Parent/Primary Caregiver Interview.** Once again, the parent interviews were conducted in the child's home with a parent or primary caregiver living with and responsible for raising the child. The interviews were conducted in the parent's language with English and Spanish versions available. Parents speaking other languages were interviewed with the aid of an interpreter. Some topics added for the spring included the child's transition from preschool to kindergarten and any information on services the family received to assist with this transition.
- **Child Assessments.** In spring 2003, the same fall assessment battery was administered with the addition of a Letter Naming Task. Children previously assessed in Spanish were assessed in English, but these children were also administered two Spanish language tests (see Exhibit 1.2).
- **Care Setting Observation.** Direct observations of care setting and quality were used for children in center-based and family day care home programs, including those participating in Head Start. These tools provide direct measures of the extent to which Head Start centers, and other childcare programs, employ skilled teachers and provide developmentally appropriate environments and curricula for their pupils. Trained observers conducted observations in classrooms and centers attended by the sampled children. Observers spent enough time in each class to ensure observation of a major portion of the daily schedule and a variety of classroom and center activities.

  The observers used standardized observational methods and coding schemes that have been widely used in child development research and whose utility has been proven in previous large-scale studies. These include: the Early Childhood Environment Rating Scale (revised) (ECERS-R), the Classroom Observation of Teacher-Directed Activities Checklist, the Arnett Scale of Teacher/Provider Behavior, and the related Family Day Care Rating Scale (FDCRS) for observations in non-center-based settings. In the interest of having some comparable observational measure of quality across all settings, a five-question observational instrument designed for use in formal care settings as well as in the home was developed. These items are completed by interviewers regardless of whether the child's care setting is Head Start, formal child care, or at home with a parent or other care provider. Assessors rate all care settings in five areas similar to the areas observed with the ECERS-R and FDCRS: overall safety, basic hygiene standards, availability of educational materials, and overall positive and negative interaction between provider and child.

- **Teacher Surveys and Care Provider Interviews.** Teachers and other care providers are asked to describe themselves, the nature of the care setting in which they work, and the types of services they provide to the selected study children. This includes biographical information such as education and years of experience, characteristics of the center or child care program, quality of program management, and belief scales to assess staff attitudes about working with and teaching children. Items on the use of literacy-promoting activities are included in the teacher survey and care provider interviews, as well as in the center director interview (see below). An "other-care provider" interview is used to collect comparable information regarding child care for non-Head Start children who were in non-center based settings or at home with a relative or non-relative (other than the parent or primary caregiver[15]). This interview includes questions on the number of children in the care setting, types of child activities used, beliefs on how children should be taught and managed, options for parent and family involvement, staffing, and respondent demographic information.
- **Teacher's/Care Provider's Child Reports (TCRs)**. Teachers and other care providers are also asked to rate each of the children in their classroom or care who are participating in the study. The following scales are used: teacher/provider relationship with child, classroom behavior and conduct, problem solving and initiative, social relationships, creative representation, music and movement, language ability, and mathematical ability. Parent and teacher/other care provider ratings of children's accomplishments and behavior are obviously not as objective as direct assessments or observations by impartial observers. Nevertheless, such ratings are an important source of information about children's learning and behavior because parents, teachers, and care providers see children over extended periods of time and in a variety of settings, providing for more robust appraisals of children's skills and competence.
- **Center Director Setting Interviews**. This in-person interview is used to collect information on the operation and quality of Head Start and non-Head Start center-based programs. Issues addressed in this interview include: staffing and recruitment, teacher education initiatives and staff training, parent involvement, curriculum, classroom activities and assessment, home visits, kindergarten transition, and demographic information about the director.

[15] Some questions from this interview were also added to the parent/primary caregiver interview to obtain comparable quality of care information for children whose care settings are their own homes.

Future Data Collection

In subsequent years, the fall telephone (and, where necessary, in-person) parent updates will be continued to obtain critical contact and care setting information. In the spring, through the child's 1st grade year, in-person parent interviews will be conducted as will the direct child assessments, with the test battery modified for the kindergarten and 1st grade years. However, once the child is in elementary school, all assessments will be conducted in the child's home.

The second spring data collection (i.e., spring 2004) has been completed, but the data have not yet been analyzed. Consequently, this report only focuses on the findings from the first study year. For the kindergarten battery, three additional Woodcock-Johnson III subtests were added: Passage Comprehension, Word Attack, and Quantitative Concepts (Concepts and Number Series). In addition, a Writing Name task has been added, and the McCarthy Draw-A-Design, Color Names & Counting, Comprehensive Test of Phonological and Print Processing (CTOPPP) Print Awareness and Story & Print Concepts have been deleted from the kindergarten battery.

Classroom observations are not being conducted in the elementary schools, but information will be collected in the spring from classroom teachers and other care providers by asking them to complete self-administered surveys and teacher reports on individual children. During the kindergarten year, the teacher survey obtains information about the kindergarten program, provisions that are made for the child's transition to kindergarten, and whether the teacher obtained any information from the Head Start program or alternative care provider about the child's development status or special needs. Also, at the kindergarten and 1st grade level, school-level data will be obtained by linking schools attended by study children to annual data collected from every public school in the U.S. by the Department of Education's National Center for Education Statistics (NCES). This includes the Common Core of Data for Public and Private Elementary Schools (CCD), and the Schools and Staffing Survey—Data for Public and Private Elementary Schools. We also plan to augment the NCES data by linking to state- and district-level data that are publicly available as school "report cards" under state accountability systems.

Response Rates

As shown in Exhibit 1.4, the individual response rates for both child assessments and parent interviews completed for the two data collection periods addressed in this report have been very good. Overall, 83 percent of parents completed interviews at both points in time, and 82 percent of the children were assessed. There is some difference between Head Start and non-Head Start groups, but the gap was slightly narrowed by the spring 2003 interview.[16]

Exhibit 1.4: Comparison of Response Rates for Head Start and Non-Head Start Groups for Fall 2002 and Spring 2003

Instrument	Fall 2002		Spring 2003	
	Head Start (Treatment)	Non-Head Start (Control)	Head Start (Treatment)	Non-Head Start (Control)
Child Assessments	85%	72%	88%	77%
Parent Interview	89%	81%	86%	79%

These response rates represent the actual (unweighted) number of interviews completed, i.e., the percentage of the sampled population that completed the interview or assessment. However, the **weighted** response rates are best for assessing the potential for nonresponse bias. The various levels of sampling where nonresponse can occur are: (1) nonresponding programs in fall 2002; (2) nonresponding centers in fall 2002; (3) additional nonresponding programs in spring 2003; (4) additional nonresponding centers in spring 2003; and (5) nonresponding children/parents in spring 2003. The overall response rate for impact estimates in the spring is the product of the response rates for each of these five levels.

Response rates of 100 percent were achieved among programs in both fall and spring, and there were no additional nonresponding centers in spring 2003. The fall center response rate was 98.9 percent, and the weighted spring child assessment response rate was 86.9 percent for the Head Start group and 76.5 percent for the non-Head Start group. Therefore, the product of the response rate for these five levels for child assessments is 80.9 percent (85.9% for the Head Start group and 75.6% for the non-Head Start group). The overall weighted response rate for parent interviews was nearly identical at 81.0 percent (86.9% for the Head Start group and 76.5% for the non-Head Start group).

[16] A high response rate has been maintained in subsequent data collection efforts. In fall 2003, 87 percent of the Head Start group and 79 percent of the non-Head Start group were interviewed. For spring 2004, 84 percent of the Head Start parents and children and 76 percent of the non-Head Start parents and children were interviewed and/or assessed.

Although the response rate is relatively high, bias in estimates of the impact of Head Start can occur to the extent that the impact differs between responding and nonresponding centers and between responding and nonresponding children and parents. As part of the weighting procedure, separate nonresponse adjustment factors were applied for categories of centers and children (see Appendix 1.2 for details). To the extent that nonrespondents and respondents within a category have similar impacts from Head Start, the application of these adjustment factors reduces the bias due to nonresponse.

Contents of This Report

This report is a preliminary examination of the impact of Head Start for children who applied to Head Start in 2002. It includes a subset of the first year of child and parenting practice outcomes as of spring 2003. This is just the precursor to the wealth of information that this study will eventually provide. This report is two volumes. In this volume, this chapter (Chapter 1) presented the study background, including an overview of the study objectives, sample design, data collection, and response rates. Chapter 2 provides further details about the study sample, including a description of child and parent characteristics measured before and after random assignment. In order to provide a context in which to understand the impact findings, Chapter 3 provides a discussion of the impact of Head Start on the types of preschool and child care settings that parents selected for their children as well as descriptive information on the characteristics of different types of early care arrangements. Chapter 4 presents an overview of the methods used for analyzing impacts on children and families.

The remaining chapters present the results of the impact analyses. The impact of Head Start on children's cognitive development is presented in Chapter 5, focusing on six different cognitive constructs, i.e., pre-reading, pre-writing, vocabulary, oral comprehension and phonological awareness, early math skills, and parent reports of children's literacy skills. The impact of Head Start on social-emotional development is presented in Chapter 6, focusing on parent reported measures of social competencies, social skills and positive approaches to learning, and problem behaviors. Chapter 7 presents findings on the impact of Head Start on children's health status and access to health care. Chapter 8 presents the findings on the impact of Head Start on parenting practices in the areas of educational activities, discipline strategies, and child safety practices. In addition, a number of technical appendices present further details about the study design, study sample, and analytic techniques.

Chapter 2: Description of the Study Sample

Representing the National Head Start Population

Because this study, as discussed in Chapter 1, is a national probability sample of Head Start programs, an important question is "Can the study findings be generalized to the complete Head Start population?" For this purpose, the population, or universe, of interest is all newly entering 3- and 4-year-olds in all Head Start centers operating in 2002-03, except those serving only special populations (i.e., programs serving primarily **only** migrant, Native American, or Early Head Start children), or very new centers (see Chapter 1 for details). Ideally, all such children would have the possibility of being included in the study and the "coverage rate" would, therefore, be equal to 100 percent.

The major cause for any undercoverage is the ethical design constraint adopted that required that the selected Head Start grantees/delegate agencies and centers have more eligible applicants than could be served at their current Federal funding level. Programs that were serving essentially all the eligible children in the community (referred to as a "saturated" program or center) could not be included in the study because including them could have resulted in a reduction in the number of children being served by Head Start.

As noted in Chapter 1, there were four points in the sample selection process where grantees/delegate agencies or centers were lost due to such saturation. First, some Head Start grantees/delegate agencies were determined to be saturated before the sample was selected, and these programs were, therefore, dropped from the sampling frame. Second, after the initial sample of grantees/delegate agencies was selected, some additional programs were found to be saturated and were also deleted from the sample. At this same point in the process, two additional programs were dropped from the sample because they were Head Start Quality Research Centers (QRC)[1] and were deleted so as not to be overburdened. The third point at which saturated sites were dropped from the sample was during the selection of Head Start centers. As with grantees/delegate agencies, some centers were initially determined to be saturated and were considered to be ineligible for inclusion and deleted from the study sample. Some centers were determined to be saturated during later attempts to conduct random assignment and also had to be dropped from the study sample.

---

[1] The Head Start Bureau and the Office of Program, Research and Evaluation (OPRE) of DHHS awarded eight cooperative agreements under the Head Start Quality Research Center (QRC) Consortium II (2001-06) to study promoting approaches to the school readiness of Head Start children.

Taking into account all of these opportunities for Head Start grantees/delegate and centers to be deleted due to saturation (or being a QRC site), the estimated weighted national coverage rate[2] for spring 2003 data was 84.5 percent, i.e., the study sample is representative of 84.5 percent of the total universe of all newly entering 3- and 4-year-olds across the country. The weight that is used for this estimate accounts for the probability of selection for each program and center and also weights the contribution of programs and centers according to the size of their enrollment. (The small number of grantees/delegate agencies and centers that were found to be closed or merged into another program or center were properly considered as ineligible, not as undercovered.)

In addition to these fully saturated grantees/delegate agencies and centers, a number of sampled centers were found to be "partially saturated," that is, there were enough applicants at the center to permit some children to be assigned to the control group, but the number available was insufficient to allow the selection of the full targeted sample. In such situations, treatment and control group children were selected from a "reserve" center, or/and a larger sample size was selected from another sampled center (in the same geographic cluster), to make up for the shortage of study children.

As discussed in the "Random Assignment" section of Chapter 1, additional undercoverage of children occurred because grantee-requested "high-risk" children were excluded from the study. The coverage rate of 84.5 percent cited above does not account for these few exclusions. These exclusions have negligible effect on the overall coverage rate, however, as there were only 276 exclusions out of approximately 18,000 newly entering applications received in the targeted programs.

To account for the undercoverage attributable to these different factors, the program and center weights were ratio adjusted to the total newly entering enrollment in the PSU and program, respectively (see Appendix 1.2). A weighting adjustment was used that was based on information obtained from the Head Start National Reporting System (HSNRS). This adjustment (possible only for children in the 4-year-old group) accounts for differences between the selected sample of Head Start grantees/delegate agencies and centers and the complete national program sampling frame. Appendix 2.1 provides a comparison of the characteristics of saturated and nonsaturated programs.

---

[2] An unweighted coverage rate can also be calculated, but this is a less useful measure of coverage as it estimates the proportion of children in the *sample*, not the universe of children served by Head Start nationally who are in programs and centers that are not saturated.

The Success of Random Assignment

An equally important question to ask about the study sample is "Was random assignment implemented well enough to support the intended impact analysis?" This question is addressed below from two perspectives. First, the characteristics of children randomly assigned to the Head Start and non-Head Start groups are compared using information collected for each child at the time of random assignment. Then, the extent to which children complied with their assigned status is examined, i.e., to what extent did children assigned to the Head Start group actually receive some Head Start services?

Comparing Head Start and Non-Head Start Children at Baseline

Exhibit 2.1 provides, separately for the 3- and 4-year-old age groups, a comparison of children randomly assigned to the Head Start and non-Head Start groups using weighted data[3] on all characteristics that were measured and available at the time of random assignment. These data were drawn from parental applications for Head Start. As shown, there are no statistically significant differences between the two randomly assigned groups, indicating that they do not differ to any discernible extent. This suggests that the initial randomization was done with high integrity and that the samples can provide the necessary confidence in the validity of the impact estimates.

Although not related to the success of random assignment, it is interesting to note that the racial/ethnic characteristics of newly entering children in the 3-year-old group are noticeably different from the characteristics of children in the newly entering 4-year-old group. This difference shows that newly entering 3-year-olds are relatively evenly distributed between the Black and Hispanic groups (32.8% vs. 37.4%), while about half of newly entering 4-year-olds are Hispanic (51.6% vs. 17.5% Black). This difference for newly entering 4-year-olds is confirmed by an examination of data from the HSNRS.[4] This ethnic difference is also reflected in the age-group differences in child and parent language.

Deviations from Random Assignment

Random assignment rarely, if ever, results in perfect adherence to the assigned program status. That is, one would expect some children assigned to the Head Start group to not participate

---

[3] The weights used are the same as those used for all the analyses discussed in this report. Details are provided in Appendix 1.2.

[4] See Appendix 2.3 for details.

Exhibit 2.1: Comparison of Head Start and Non-Head Start Study Groups: Child and Family Characteristics Measured Prior to Random Assignment (Weighted Data)

Characteristic	Head Start (Treatment) Group	Non-Head Start (Control) Group	Difference: (Head Start) – (Non-Head Start)
Child Gender:			
3-Year-Old Group			
Boys	48.5%	48.9%	-0.4%
Girls	51.5%	51.1%	0.4%
4- Year-Old Group			
Boys	51.1%	49.4%	1.7%
Girls	48.9%	50.6%	-1.7%
Child Race/Ethnicity:			
3- Year-Old Group			
White	24.5%	26.6%	-2.1%
Black	32.8%	31.8%	1.1%
Hispanic	37.4%	35.7%	1.6%
Other	5.3%	5.9%	-0.6%
4- Year-Old Group			
White	26.7%	23.3%	3.4%
Black	17.5%	17.0%	0.5%
Hispanic	51.6%	53.8%	-2.1%
Other	4.1%	5.9%	-1.8%
Child Language:			
3- Year-Old Group			
English	71.1%	69.9%	1.2%
Spanish	24.8%	24.0%	0.8%
Other	3.9%	5.7%	-1.8%
Missing	0.2%	0.4%	-0.2%
4- Year-Old Group			
English	57.1%	56.4%	0.8%
Spanish	39.3%	40.8%	-1.5%
Other	3.2%	2.3%	0.8%
Missing	0.4%	0.5%	-0.1%
Parent Language:			
3- Year-Old Group			
English	74.8%	74.8%	0.0%
Spanish	23.1%	22.0%	1.1%
Other	1.5%	1.7%	-0.2%
Missing	0.6%	1.5%	-0.9%
4- Year-Old Group			
English	59.5%	58.4%	1.1%
Spanish	37.8%	39.5%	-1.7%
Other	0.9%	0.5%	0.5%
Missing	1.8%	1.6%	0.2%
Child Income Eligible:			
3- Year-Old Group			
No	7.7%	6.7%	1.0%
Yes	91.4%	91.9%	-0.6%
Missing	0.9%	1.4%	-0.5%
4- Year-Old Group			
No	6.0%	10.1%	-4.0%
Yes	91.8%	87.9%	3.9%
Missing	2.2%	2.1%	0.1%

Notes: (1) Data source: Roster information used at time of random assignment; (2) T-tests of the difference between the Head Start and non-Head Start percentage in each row were run for each characteristic; no statistically significant differences were found. With large samples, differences in means for 0/1 variables (e.g., 1=boys, 0=girls) have approximately normal distributions and follow the t distribution once divided by their standard errors.

in the program (referred to as "no-shows"), and some of the children assigned to the non-Head Start group to enroll in the program (referred to as "crossovers").

Such violations of pure random assignment were not, therefore, unexpected. During program recruitment, Head Start grantees and centers reported "no-shows" as a challenge they confront, with rates often in the double-digits. Absent a requirement that parents and children participate once they are accepted for Head Start enrollment, it is not surprising that some families who were randomly assigned to the Head Start group in the study subsequently opted for a different care setting for their child.[5]

Similarly, although every effort was made to maintain the integrity of the non-Head Start comparison group, perfect conditions could not be implemented. In a few rare instances, local staff intentionally enrolled non-Head Start children into Head Start. However, a greater threat to compliance was that parents could apply to another nearby Head Start program. This problem was particularly an issue in densely populated areas with two or more Head Start programs operating in close proximity. And, due to confidentiality restrictions, local study staff were not able to share information on participants with other nearby grantees, reducing the ability to keep control group families from Head Start enrollment.

Exhibit 2.2 provides information on the incidence of Head Start group "no-shows" and non-Head Start group "crossovers" by age group for both the total sample randomly assigned and for the children who are part of the Year 1 analysis sample that forms the basis for the findings reported in subsequent chapters of this report. The Year 1 analysis sample includes only those children (and their parents) for whom data could be collected in spring 2003 (see Chapter 4 for details on the analysis sample). In the exhibit, a child in the Head Start group is considered a "no show" if it was determined that he/she did not participate in Head Start **at any time** during the 2002-03 program year. A child in the non-Head Start group was deemed a "crossover" if he/she participated in Head Start **at any time** during the 2002-03 program year. This determination (explained in more detail in Appendix 2.2) was based on information obtained from parent surveys in fall 2002 and spring 2003, follow-back contact with all Head Start centers in the study in fall 2002 to see if individual children had attended Head Start, and care setting identified at the time of the child's fall 2002 and spring 2003 assessments.

---

[5] Chapter 3 presents a breakdown of the types of settings children attended.

As shown in this table, “no-shows” accounted for 15 and 20 percent of the full randomly assigned sample for children in the 3- and 4-year-old groups, respectively, and 12 and 17 percent of the Year 1 analysis sample for 3- and 4-year-old groups once analysis weights were applied. Similarly, crossovers accounted for 17 and 14 percent of the randomly assigned group, and 19 and 17 percent of the analysis sample. The resulting differences across the two samples—a lower incidence of “no-shows” and a higher incidence of “crossovers” in the analysis sample compared to all randomly assigned children—are probably due to higher response rates among children in Head Start programs (i.e., they were probably easier to find).

Exhibit 2.2: The Incidence of No-Show and Crossover Behavior for Both the Sample as Randomly Assigned and the Year 1 Analysis Sample, by Age Cohort (Weighted Data)

Sample Group	Any Year 1 Head Start Participation	No Year 1 Head Start Participation	Total
All Randomly Assigned (N=4,667):			
3-Year-Old Group			
Head Start Group	85.1%	14.9%	100%
Non-Head Start Group	17.3%	82.7%	100%
4-Year-Old Group			
Head Start Group	79.8%	20.2%	100%
Non-Head Start Group	13.9%	86.1%	100%
Year 1 Analysis Sample (N=3,898):			
3-Year-Old Group			
Head Start Group	88.2%	11.8%	100%
Non-Head Start Group	18.5%	81.5%	100%
4-Year-Old Group			
Head Start Group	83.4%	16.6%	100%
Non-Head Start Group	16.5%	83.5%	100%

Chapter 4 explains how impact estimates are adjusted to account for these occurrences. At this point, it is important to note that the observed levels of noncompliance with the design, although not to be dismissed, are not atypical of what has been found in other random assignment studies and do not undermine the basic validity of the study. At worst, violations of random assignment that extend Head Start’s services to some children in the non-Head Start group and reduce the exposure to Head Start among the treatment group make it harder to detect any average impact of the program that does occur with the available sample size. These considerations should increase the confidence that any observed statistically significant impacts are real and important. The downside, of course, is that some effects of Head Start may be obscured by the loss of analytic power due to the presence of “no-shows” and “crossovers.”

Characteristics of the Year 1 Study Sample

This final section of Chapter 2 examines the characteristics of the current analysis sample that is used in this report. It comprises those children, and their parents, from whom data were collected in spring 2003. Exhibits 2.3-A and 2.3-B compare the characteristics of the children between the Head Start and non-Head Start groups for the 3- and 4-year-old groups, respectively. (The figures in these tables differ from those in Exhibit 2.1, which looked at **all** children randomly assigned.) The characteristics for these comparisons were all measured in fall 2002 and ideally represent the "baseline" or pre-intervention point of the study. They are also the characteristics, as discussed in Chapter 4, used as covariates in the statistical models to estimate program impacts, or to examine any variation in program impacts for particular population subgroups (e.g., are impacts higher or lower for children with disabilities?).

As demonstrated by these two tables, the Head Start and non-Head Start groups do not differ to any discernible extent except for two small differences for each of the two age groups. For the 3-year-old group, the primary caregivers of children in the Head Start group are slightly older than caregivers of children in the non-Head Start group; and the Head Start group children are somewhat more likely to have a grandparent living with them. In the case of the children in the 4-year-old group, the mothers of children in the Head Start group are more likely to have attained an educational level beyond high school, and the households of children in the Head Start group are somewhat less likely to receive public assistance through the Federal TANF program. As discussed in Chapter 4, these differences may arise from the lag in fall 2002 data collection after the point of random assignment. Because the differences are not fully accounted for by the nonresponse adjustments to the sampling weights (see Appendix 1.2), they are included in the analysis as covariates in the statistical models used to estimate program impacts (see Chapter 4).

Exhibit 2.3-A: Description of the Year 1 Analysis Sample: 3-Year-Old Group (Weighted Data)

Characteristic	Head Start (Treatment) Group	Non-Head Start (Control) Group	Difference: (Head Start) – (Non-Head Start)
Child Gender:			
Boy	47.9%	49.1%	-1.2%
Girl	52.1%	50.9%	1.2%
Child Race/Ethnicity:			
White	24.3%	26.0%	-1.7%
Black	33.3%	31.4%	1.9%
Hispanic	37.0%	36.4%	0.6%
Other	5.4%	6.3%	-0.8%
Child Has a Disability	13.5%	11.9%	1.6%
Fall-Spring Language of Child Assessment:			
English-English	75.4%	75.9%	-0.5%
Spanish-English	18.9%	18.0%	0.9%
Spanish-Spanish	4.3%	4.6%	-0.3%
Primary Home Language Is English	71.9%	68.5%	3.4%
Biological Mother Was a Teen Mom	36.2%	37.6%	-1.3%
Biological Mother Is a Recent Immigrant	17.0%	17.8%	-0.8%
Biological Mother Is Employed	51.4%	57.4%	-6.0%
Both Biological Parents Live With Child	48.5%	50.7%	-2.2%
Child's Parents Are			
Married	43.7%	45.3%	-1.6%
Separated or Divorced	11.5%	13.7%	-2.2%
Primary Caregiver's Age as of 9/1/02	29.5 years	28.6 years	**0.9 years***
Mother's Education:			
Less Than High School	32.4%	34.8%	-2.3%
High School/GED	34.7%	33.9%	0.8%
Beyond High School	32.9%	31.4%	1.5%
Grandparent Lives in Home	3.6%	1.7%	**1.9%****
Parent's Self-Reported Health Is Excellent or Good	85.5%	86.5%	-1.0%
Primary Caregiver—Depression Scale	251.9	251.2	0.7
Primary Caregiver—Locus of Control Scale	249.5	251.2	-1.7
Average Household Income:			
$500/month or less	14.8%	12.0%	2.9%
$501-$1500/month	48.3%	53.4%	-5.1%
Over $1500/month	36.9%	34.6%	2.3%
Household Receives TANF	10.6%	10.5%	0.1%

*= p≤0.05, ** = p≤0.01, *** = p≤0.001.
Data source: Roster information collected at the time of random assignment and fall 2002 Parent Survey.

Exhibit 2.3-B: Description of the Year 1 Analysis Sample: 4-Year-Old Group (Weighted Data)

Characteristic	Head Start (Treatment) Group	Non-Head Start (Control) Group	Difference: (Head Start) – (Non-Head Start)
Child Gender:			
Boy	49.6%	51.2%	-1.6
Girl	50.4%	48.8%	1.6
Child Race/Ethnicity:			
White	27.8%	24.6%	3.2%
Black	25.5%	23.3%	2.2%
Hispanic	42.4%	45.8%	-3.4%
Other	4.3%	6.2%	-1.9%
Child Has a Disability	12.8%	11.4%	1.4%
Fall-Spring Language of Child Assessment:			
English-English	67.2%	64.3%	2.9%
Spanish-English	25.9%	28.3%	-2.5%
Spanish-Spanish	5.9%	5.4%	0.4%
Primary Home Language Is English	63.6%	63.2%	0.0%
Biological Mother Was a Teen Mom	38.6%	35.2%	3.4%
Biological Mother Is a Recent Immigrant	24.1%	23.5%	0.6%
Biological Mother Is Employed	48.5%	52.0%	-3.4%
Both Biological Parents Live with Child	51.3%	51.3%	0.0%
Child's Parents Are			
Married	45.2%	45.4%	-0.2%
Separated or Divorced	15.9%	14.9%	1.0%
Primary Caregiver's Age as of 9/1/02	29.3 years	29.5 years	-0.2 years
Mother's Education:			
Less Than High School	38.6%	41.6%	-3.0%
High School/GED	31.7%	35.2%	-3.5%
Beyond High School	29.8%	23.3%	**6.5%***
Grandparent Lives in Home	2.4%	1.4%	1.0%
Parent's Self-Reported Health Is Excellent or Good	86.6%	86.4%	0.1%
Primary Caregiver—Depression Scale	251.1	248.6	2.4
Primary Caregiver—Locus of Control Scale	251.5	249.5	2.0
Average Household Income:			
$500/month or less	11.8%	9.1%	2.7%
$501-$1500/month	46.2%	50.8%	-4.6%
Over $1500/month	42.0%	40.0%	2.0%
Household Receives TANF	10.0%	14.4%	**-4.5%***

*= p≤0.05, ** = p≤0.01, *** = p≤0.001.
Data source: Roster information collected at the time of random assignment and fall 2002 Parent Survey.

Chapter 3: Children's Experiences

This chapter has two purposes: (1) to present findings of the impact of Head Start on the types of preschool and child care arrangements that parents select for their children and (2) to present descriptive information (not program impacts) on the characteristics of different types of early care arrangements used for children. The descriptive section focuses on a comparison between the Head Start and other center-based classrooms that study children attended and uses observational data (the ECERS-R and Arnett Scale of Teacher Behavior) and survey data on the reported types of activities teachers did with children and the types of curricula they used in their classrooms. Together these descriptions provide a context for understanding the impact findings presented in subsequent chapters.[1]

Highlights

Impact Findings

- The importance of Head Start as an early care option for low-income families is demonstrated by comparing the care arrangements used by parents of children in the Head Start and non-Head Start groups. Providing children with access to Head Start had a statistically significant impact on children's use of parent care and center-based care. Specifically:
    1. Non-Head Start group children were substantially more likely than the Head Start group children to be in parent care in spring 2003. Children were considered in parent care if they did not have a preschool or child care arrangement for at least 5 hours per week. Among children in the 3-year-old group, 39.2 percent of non-Head Start group children were in parent care as compared to only 6.8 percent of children in the Head Start group. Among the 4-year-old group, the figures were 41.6 and 8.7 percent, respectively.
    2. Head Start group children were twice as likely as the non-Head Start group children to use a center-based program in spring 2003. Approximately 90 percent of children in the Head Start group in both age groups were using a center-based program compared to 43 percent of children in the 3-year-old non-Head Start group and 48 percent of the 4-year-old group.
    3. The child care arrangements used for children in the study did not differ substantially by age group. This is a somewhat surprising finding given that research about the use of center-based programs by 3- and 4-year-olds in population-based samples tends to show that 4-year-olds are more likely than younger children to be enrolled in center-based programs
    4. Head Start group children were more likely than non-Head Start group children to be in a center-based environment in both the fall 2002 and spring 2003 and to have been in their spring 2003 setting since the start of the 2002-03 program year.

[1] The current analyses of quality provide only a comparison of Head Start and non-Head Start centers and do not look at how impacts vary as a function of varying levels of quality. Future analyses will address this issue.

Descriptive Findings

- These findings focus on some initial quality indicators for the Head Start centers and other center-based programs attended by study children. On the initial indicators assessed, children in the Head Start centers were in environments that more often had positive interactions between children and teachers, used curriculum and activities to enhance children's skills, and had higher scores on the ECERS-R. Specifically:

  1. Children in the 3-year-old group in Head Start classrooms were significantly more likely to be in classrooms with higher average total scores than children in other center-based programs. The average ECERS-R score was 5.17 as compared to 4.44 for the other center-based programs. Similarly for children in the 4-year-old group, the mean ECERS-R score in Head Start classrooms was 5.29 as compared to 4.62 for the other center-based classrooms.
  2. As measured by the Arnett Scale of Teacher Behavior, children in the 3-year-old group in Head Start Centers had teachers who were rated as more sensitive and who promoted more independence than the teachers of children in other center-based programs. For the children in the 4-year-old group, Head Start teachers were rated to be less harsh and to have promoted more independence than those in other center-based programs.
  3. Teachers of children in the 3-year-old group in Head Start classrooms reported using language and literacy activities more frequently than teachers in other center-based classrooms (almost half of the 11 literacy activities). No significant difference was found in the frequency of these activities for children in the 4-year-old group.
  4. Teachers of children in Head Start classrooms (for both age groups) reported conducting math activities more frequently than teachers in other center-based classrooms.
  5. Children enrolled in Head Start were more likely to have teachers who used a curriculum. For children in other center-based classrooms, approximately 14 percent of the 3-year old group and 17 percent of the 4-year-old group were in classrooms that did not use a curriculum. This compares to about 2 percent of the children in the 3-year-old group and 4 percent of the children in the 4-year-old group who were in Head Start classrooms.
  6. In classrooms where a curriculum was used, there was more uniformity in the type of curriculum used by Head Start classrooms as compared to other center-based classrooms. More than three-fourths of children in Head Start classrooms were in classrooms using High Scope or Creative Curriculum compared to about half of the children in other center-based classrooms.

Impact on Children's Early Care Settings

The findings in this section describe the impact of having access to Head Start on the preschool and child care arrangements used by low-income families that apply to, and are eligible for, Head Start. Specifically, the results highlight the extent to which families who have access to Head Start are actually enrolled at various points in time and what early care services they used when they did not gain access to Head Start services.

As discussed in Chapter 1, the parents of children in the control group were not precluded from enrolling their children in other types of preschool or child care arrangements. Consequently, the impact of Head Start is being evaluated against a mixture of alternatives available in the community, ranging from parent care to center-based programs as defined in Exhibit 3.1. In some cases, these alternative arrangements may look very much like Head Start in their characteristics, while others may look very different from Head Start. Understanding the extent to which children in the control group children use various alternatives is, therefore, vital for understanding the services to which Head Start is being compared and for interpreting the estimates of Head Start's impact.

Exhibit 3.1: Definition of Children's Preschool and Child Care Arrangements

Types of Care Arrangements

Head Start: center-based, home-based, and combination programs funded with Federal Head Start dollars. Children in center-based programs with a mix of funding sources were placed into this category if they were enrolled in a classroom that received any Federal Head Start dollars.

Non-Head Start Center: center-based programs as differentiated from care that takes place in someone's home or federally funded Head Start programs. Some children in this category are enrolled in centers that receive Federal Head Start dollars but are not in classrooms that receive any Federal Head Start dollars.

Relative's and Non-Relative's Home: non-parental care that takes place in a home that is not the child's own home, either by a relative or a non-relative of the child. This category includes regulated family child care providers as well as home-based child care providers who are exempt from regulation by state and/or local licensing agencies.

Child's Home with a Relative or Non-Relative: non-parental care that takes place in the child's own home, either by a relative or a non-relative of the child. Caregivers in this category are typically not subject to regulation by state and/or local licensing agencies.

Parent Care: care by the child's parent or guardian, typically in the child's own home.

Definition of Focal Arrangement

A child's focal arrangement is defined as either the treatment or primary alternative to the treatment. Head Start is always defined as the focal arrangement for children enrolled in Head Start. For all other children, the focal arrangement is generally defined as the non-parental arrangement (if there was one) the child attended between the hours of 8 AM and 6 PM Monday through Friday, for at least 5 hours per week. For children in multiple arrangements that met these criteria, the following hierarchy was used to prioritize and select a setting: center-based programs, followed by non-relative's homes, relative's homes, and finally care by a non-parental relative in the child's home. In the absence of non-parental care that meets the time criteria, the child's focal setting is parent care.

Because there is reason to believe that families applying to Head Start may be different from the overall population of low-income families, at least in terms of motivation to enroll their children in a preschool program prior to kindergarten, the data do not speak to the arrangements used by low-income families more generally; they speak only to those families who applied to Head Start. Additionally, the Head Start centers are nationally representative of Head Start. In contrast, the non-Head Start centers are not nationally representative, but instead represent the types of center-based care families use when their children could not go to Head Start.

Data and Methods

The results reported in this section rely on data collected through parent interviews completed in fall 2002 and spring 2003. Parents were asked where their child regularly spends time Monday through Friday, who was responsible for him/her during that time period, and the start dates of the reported arrangements. Head Start children may attend other programs during the hours they are not in Head Start and, as a consequence, may be exposed to numerous other experiences at home and elsewhere that shape their development. Similarly, children assigned to the non-Head Start group also may have a variety of experiences that affect their development, such as care provided in their own home, as well as time spent in other settings.

Although all care experiences are important in understanding children's developmental outcomes, practical considerations and differences in the nature of the settings limit the type and depth of data that can be collected across these various arrangements. Consequently, criteria were developed to help identify, categorize, and prioritize the range of settings in which children spend time and to identify a focal arrangement (see Exhibit 3.1). For children attending Head Start, the focal arrangement is defined as Head Start. For children not attending Head Start, the focal arrangement is generally defined as the non-parental arrangement (if there was one) attended for at least 5 hours per week, at least in part between the hours of 8:00 AM and 6:00 PM, Monday through Friday. If children participated in multiple non-parental arrangements, a hierarchy was used to prioritize and select the focal arrangement as follows: center-based programs, followed by non-relatives' homes, relatives' homes, and care by someone other than the parent in the child's home. In the absence of any non-parental arrangement, the focal arrangement is parental care.

This definition of children's focal arrangements ensured consideration of arrangements that, though not necessarily occupying the majority of the child's time, may well affect the child's development by offering at least some of the supplementary educational, social, and access to service opportunities offered by Head Start. We also explored an alternative definition of the focal arrangement that was based on where children spent the most time between 9 AM and 3 PM. Differences in results across the definitions were minimal.[2] Furthermore, expanding the definition to capture non-parental settings of at least 5 hours provides the opportunity to capture the effect that preschool and child care services might have on the development of each group. Home environment influence is captured through the parent interview. The findings in this section are based on the child's focal care setting. Observations and teacher input were obtained at the child's focal setting.

Impact on Children's Spring 2003 Focal Arrangements

Exhibit 3.2 presents, by age group, the percentage of children in the Head Start and non-Head Start groups who were in each of seven types of focal arrangements at the time of the spring 2003 data collection. As shown, providing children with access to Head Start had a statistically significant impact on the preschool and child care arrangements used by children in the study.[3] These differences are most notable in terms of the different rates at which families rely on parent care and center-based programs.

Among children in the 3-year-old group, 39 percent of the children in the non-Head Start group used parent care as the primary form of care compared to only 6.8 percent of the children in the Head Start group. Similarly, among children in the 4-year-old group, the figures were 41.6 and 8.7 percent, respectively. Approximately 90 percent of Head Start group families (for both age groups) were using some type of center-based care (including Head Start).[4] In contrast, among non-Head Start group families, only 43 percent of the 3-year-old group and 48 percent of the 4-year-old group were in some form of center-based care.

All of the parents of the study population were interested in having their children attend Head Start. Yet, when the study created an "alternative world" in which Head Start was not available to them, two out of five non-Head Start group families kept their children at home with a parent. About the same fraction of these families enrolled their child in a non-Head Start center-based program. The remaining children in the control group were found in care in their own, or someone else's, home with an individual other than the parent. As also shown in this table, and as discussed in Chapter 2, some children assigned to the Head Start group did not attend Head Start, and some children assigned to the non-Head Start group managed to gain entrance into the program. (Note: the figures shown here are as of spring 2003 only and, as a consequence, differ slightly from data reported in Chapter 2.)

---

[2] Appendix 3.1 provides an exhibit showing the percentage of children in Head Start and non-Head Start groups by main care arrangement.

[3] Significance tests used in this exhibit, as well as Exhibits 3.3 and 3.4, are based on two-tailed t-tests. This approximates the mean of a 0/1 variable in a large sample, as described in Chapter 2.

[4] The term "non-Head Start center" is used for convenience because this category generally represents preschool and child care programs that cannot be classified as Head Start. However, Head Start services have been defined somewhat more narrowly for the purpose of this study. As a result, the category of "non-Head Start centers" actually includes some children in centers that meet Head Start Performance Standards, but these children were not enrolled in *classrooms* receiving Federal Head Start funding.

Exhibit 3.2: Percentage of Children in Head Start and Non-Head Start Groups by Type of Focal Arrangement in Spring 2003 (Weighted Data)

Type of Focal Arrangement	**Head Start Group (Sample Size=1,336)**	**Non-Head Start Group (Sample Size=821)**	**Head Start Group (Sample Size=1,068)**	**Non-Head Start Group (Sample Size=662)**
	Percentage of 3-year-old group		**Percentage of 4-year-old group**	
Parental Care	**6.8*** **	**39.2**	**8.7*** **	**41.6**
Non-Parental Care	**93.2*** **	**60.8**	**91.3*** **	**58.4**
Head Start	84.1***	17.5	76.4***	13.4
Non-Head Start center	6.9***	25.0	11.9***	34.7
Non-relative's home	0.7***	6.0	1.4*	5.0
Relative's home	1.0***	8.7	0.9	3.1
Child's home w/relative	0.6**	3.5	0.5*	2.2
Child's home w/non-relative	0.0	0.1	0.2	0.1
Total percent	100%	100%	100%	100%

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.

These findings emphasize that the impact of Head Start is being evaluated against a mixture of alternative care settings rather than against a purely "no-services" condition. All types of alternatives, including parent care, may offer an environment that effectively supports children's development. However, parent care and center-based programs may generally be thought of as falling on opposite ends of a continuum in terms of the likelihood that the environment delivers a set of services and experiences that is similar to Head Start. Children in the non-Head Start group tend to be concentrated at the two ends of this continuum, with a much smaller share in the center-based category than is true of children given access to Head Start.

Also, the child care arrangements for children in the non-Head Start group do not differ substantially between the two age groups. This is a somewhat surprising finding, given that research about the use of center-based programs by 3- and 4-year-olds in population-based samples tends to show that 4-year-olds are more likely than younger children to be enrolled in center-based programs.[5] Furthermore, public funding for pre-kindergarten and preschool programs is often targeted at 4-year-olds, which would suggest that low-income parents of 4-year-olds might find it easier to access center-based services than parents of 3-year-olds. As a result, it was hypothesized that 3-year-olds assigned to the non-Head Start group would use center-based programs at a lower rate than 4-year-olds assigned to the non-

[5] See: (1) US Department of Education, National Center for Education Statistics. (2002). ***The Condition of Education 2002.*** Washington, DC: US Government Printing Office (NCES Publication No. 2002-025.; and (2) The Urban Institute, (2003). ***Percentage of Three- and Four-Year Olds in Poverty in Different Types of Child Care Arrangements.*** Unpublished calculations based on data from the 1999 National Survey of America's Families.

Head Start group. However, parents of children in the 3-year-old group who want, but cannot access, Head Start tend to use other types of center-based care at roughly the same rate as parents of children in the 4-year-old group.

Exposure to Head Start or Center-Based Care in Fall 2002 and/or Spring 2003

This section extends the analysis of arrangements used in spring 2003 to consider whether children were enrolled in Head Start or another center-based program, in either fall 2002, spring 2003, or both time points. This analysis estimates how many children assigned to the Head Start and non-Head Start groups might have been exposed, over the entire course of the 2002-03 program year, to a program that may have offered the types of educational, social, and access-to-services opportunities that are offered by Head Start. In addition, this analysis begins to explore, in a general sense, **how much** exposure children had to these types of preschool opportunities by examining whether children were enrolled in these types of programs during at least two points in time.

Considering fall 2002 and spring 2003, Exhibit 3.3 shows that Head Start group children were significantly more likely than non-Head Start group children to be enrolled in Head Start or another center-based program in one or both points in time. Only 4 percent of the Head Start group children were **not** enrolled in either Head Start or another center-based program in fall 2002 and/or spring 2003. In contrast, a much larger proportion (47 percent of the 3-year-olds and 40 percent of the 4-year-olds) of the non-Head Start group children were **not** enrolled in some type of center-based program in fall 2002 and/or spring 2003.

Exhibit 3.3: Percentage of Children in Head Start and Non-Head Start Groups by Age Group and Type of Arrangement Attended in Fall 2002 and/or Spring 2003 (Weighted Data)

Type of Center Attended	Head Start Group (Sample Size =1,333)	Non-Head Start Group (Sample Size=769)	Head Start Group (Sample Size =1,047)	Non-Head Start Group (Sample Size =623)
	Percentage of 3-year-old group		Percentage of 4-year-old group	
Attended Head Start	89.4***	21.3	85.6***	18.1
Did not attend Head Start but attended a center-based program	6.2***	31.9	10.8***	42.5
No center-based arrangement attended	4.4***	46.8	3.6***	39.4
Total percent	100%	100%	100%	100%

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Furthermore, children in the Head Start group were more likely than children in the non-Head Start group to be in Head Start or another center-based program in **both** the fall and the spring. Among the Head Start group children who attended Head Start or a center-based program in fall or spring, over 90 percent of both age groups were in one of these types of programs at both points in time (see Exhibit 3.4). In contrast, only 40 percent of the non-Head Start group children were in one of these types of programs in **both** the fall and the spring.

Exhibit 3.4: Percentage of Children Attending Head Start or a Center-Based Program in Fall 2002 and Spring 2003, in Spring 2003 Only, in Fall 2002 Only, or Not at All by Head Start Group and Non-Head Start Group (Weighted Data)

Attended Head Start or a Center-Based Program	**Head Start Group (Sample Size =1,231)**	**Non-Head Start Group (Sample Size =717)**	**Head Start Group (Sample Size =999)**	**Non-Head Start Group (Sample Size =582)**
	Percentage of 3-year-old group		**Percentage of 4-year-old group**	
Fall 2002 and spring 2003	91.0***	38.3	89.8***	40.9
Fall 2002 only	3.2	5.4	5.6	7.5
Spring 2003 only	0.8***	4.1	0.6**	6.8
Neither fall 2002 nor spring 2003	4.9***	52.3	4.0***	44.9
Total percent	100%	100%	100%	100%

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Stability of Children's Settings

The previous section examined the proportion of children that attended Head Start and/or a center-based program at two points in time during the 2002-03 program year. Those results identified the length of time that children were likely exposed to specific types of preschool and child care arrangements that could affect school readiness. However, those estimates did not consider whether children were in the **same** arrangement over time. Additional insight into the length of time Head Start and non-Head Start group children were exposed to a particular preschool or child care arrangement is provided in Exhibit 3.5. This analysis is not considered an impact finding because it focuses on a subgroup of children whose childcare arrangement was affected by access to Head Start. The subgroup is those children whose focal arrangement was not parental care. For this analysis, length of time is measured by whether children had been in their spring 2003 focal arrangements since the start of the 2002-03 year. These results may also be used to gain a preliminary understanding of the stability of arrangements among children not exclusively in parental care. This is considered a preliminary understanding because the research team has not yet been able to fully explore the extent to which later start dates are a reflection of higher turnover in arrangements among certain groups of children versus a reflection of families making a single transition to preschool or Head Start later, rather than earlier, in the school year.

Exhibit 3.5: Focal Arrangement Start Dates Among Children Not Exclusively in Parent Care (Weighted Data)

Began Non-Parental Focal Arrangement	Head Start Group (Sample Size=1,222)	Non-Head Start Group (Sample Size =446)	Head Start Group (Sample Size=964)	Non-Head Start Group (Sample Size =382)
	Percentage of 3-year-olds		Percentage of 4-year-olds	
September 2002 or earlier	88.9**	77.2	90.7**	81.4
October 2002 to December 2002	6.0	7.8	4.8	8.6
January 2003 or later	5.1***	15.0	4.5*	10.0

*p≤0.05, ** p≤0.01, ***p≤0.001.

Exhibit 3.5 indicates that approximately 90 percent of children in both cohorts assigned to the Head Start group who used non-parental arrangements began their spring 2003 arrangements in September 2002 or earlier, as compared to the non-Head Start group (77 percent of the 3-year-old group and 81 percent 4-year-old group). Thus, of children using non-parental arrangements, those in the Head Start group were more likely to have been in the same setting since the beginning of the school year than those in the non-Head Start group.

As presented in this section, the majority of study children were in some type of center-based care in spring 2003. The next section provides further descriptive information about these center-based environments, focusing on some of the quality differences between Head Start centers and other center-based programs.

Description of Center-Based Classroom Environments

This section compares some initial quality indicators of the Head Start and other center-based programs that were attended by study children, without taking into account treatment and control group differences. Therefore, it provides a description of a preliminary set of quality characteristics that children experienced in these two different environments, **rather than an estimate of the impact of Head Start on the quality of care.** Although the sample of Head Start centers is nationally representative, the other center-based programs included in this analysis are not nationally representative. Instead they represent the types of center-based care families use when their children cannot go to Head Start. Future analyses will expand on the description of setting characteristics and determine how the child impacts vary with the quality of their early care experience.

Preschool programs are typically rated on two important dimensions of quality—process and structural (Phillips et al., 2000).[6] Process characteristics of the classroom environment generally include the nature of teacher-child interactions, use of curriculum, schedule of activities, and use of instructional materials. Structural indices refer to measures such as staff-child ratio, group-size, and teacher's education. As discussed in Chapter 1, this study collected information (through classroom observations, teacher surveys, and parent interviews) on a variety of setting structural and process characteristics (e.g., classroom resources, teacher-child ratio, teacher characteristics, the nature of children's every day experiences, comprehensive services provided, and parent involvement and satisfaction).

This report provides a first look at a few of the process indicators, using data from the ECERS-R, the Arnett Scale of Lead Teacher Behavior, and teacher reports on activities and curricula used in classrooms. Trained observers conducted classroom observations in spring 2003. Each classroom was visited one time, and the observers were on site for approximately 4 hours. Teachers were also given surveys and asked to self-report on a variety of elements related to teaching young children. Observations, teacher surveys and teacher reports on children were obtained from each child's focal setting. Analysis was conducted at the child level. The effect of clustering on standard errors has been accounted for through replicate weights (see Appendix 1.2 for discussion on weighting).

The Early Childhood Environment Rating Scale – Revised

The revised ECERS-R (Harms, Clifford, & Cryer, 1998)[7] is a 37-item instrument that measures a wide variety of quality-related processes occurring in the preschool classroom. It is divided into six subscales:

- **Space and Furnishings:** Eight items that rate the adequacy of the furniture and gross motor equipment and how that furniture or equipment is arranged to allow children to play, learn, relax, and have some privacy. It also rates child-related displays within the classrooms.
- **Personal Care Routines:** Six items that rate greetings/departures, meals/snacks, nap, toileting, and heath and safety practices.
- **Language and Reasoning:** Four items that rate the range of accessible books and how they are used, whether children are encouraged to communicate, use of language to develop reasoning skills, and the level of conversation between staff and children.
- **Activities:** Ten items that rate whether a variety of activities are available and used—fine motor, art, music/movement, blocks, sand/water, dramatic play, nature/science, and

[6] Phillips, D., D. Makos, S. Scarr, K. McCartney, and M. Abbott-Shim. "Within and beyond the classroom door: Assessing quality in child care centers." ***Early Childhood Research Quality***, 15 (4), 475-496.

[7] Harms, T., R.M. Clifford, and D. Cryer. (1998). ***Early Childhood Environment Rating Scale: Revised Edition***. New York: Teachers College Press.

math/number activities. This section also rates use of TV/video and computers in the classroom and the materials and activities used to promote cultural diversity.

- **Interaction.** Five items that rate the supervision of children (gross motor and general supervision), discipline used, staff-child interactions, and interaction among children.
- **Program Structure.** Four items that rate the use of a daily schedule, amount of free play and the materials provided, amount of group time, and the provisions made for children with disabilities.

Each item on the ECERS-R is given a score from 1 to 7, and items are grounded by the odd numbers with 1 = inadequate, 3 = minimal, 5 = good, and 7 = excellent care. Scores for each subscale as well as for the overall total score are reported in Exhibit 3.6 for Head Start and other center-based programs.

For children in both the 3- and 4-year-old groups, the total ECERS-R mean scores were significantly higher for the settings attended by children in Head Start classrooms than for the other center-based programs.[8] On average, the scores for children in the Head Start classrooms were in the "good" range (5.17 for the 3-year-old group and 5.29 for the 4-year-old group), while children in the other center-based programs averaged scores toward the upper end of the "minimally adequate" range (4.44 for the 3-year-old group and 4.62 for the 4-year-old group).[9]

With the exception of the interaction subscale score for the 3-year-old group, Head Start subscale mean scores were significantly better than for the other center-based programs for both age groups. As context for understanding children's language and literacy outcomes, it is particularly noteworthy that for both age groups, children in Head Start center classrooms had significantly better scores on the language and reasoning subscale than did children in the other center-based programs. A higher score indicates a richer language environment. Head Start classrooms for both age groups of children scored in the "good" range as compared to scores in the "minimally adequate" range for children in non-Head Start classrooms.

[8] Recall that these figures pool the experiences of children in both the Head Start and non-Head Start groups and represent the differences for children in Head Start classrooms versus other center-based programs, rather than impacts of the Head Start programs on quality.

[9] The standard deviation for total mean ECERS-R score for Head Start centers is 0.91 and is 1.14 for other center-based programs.

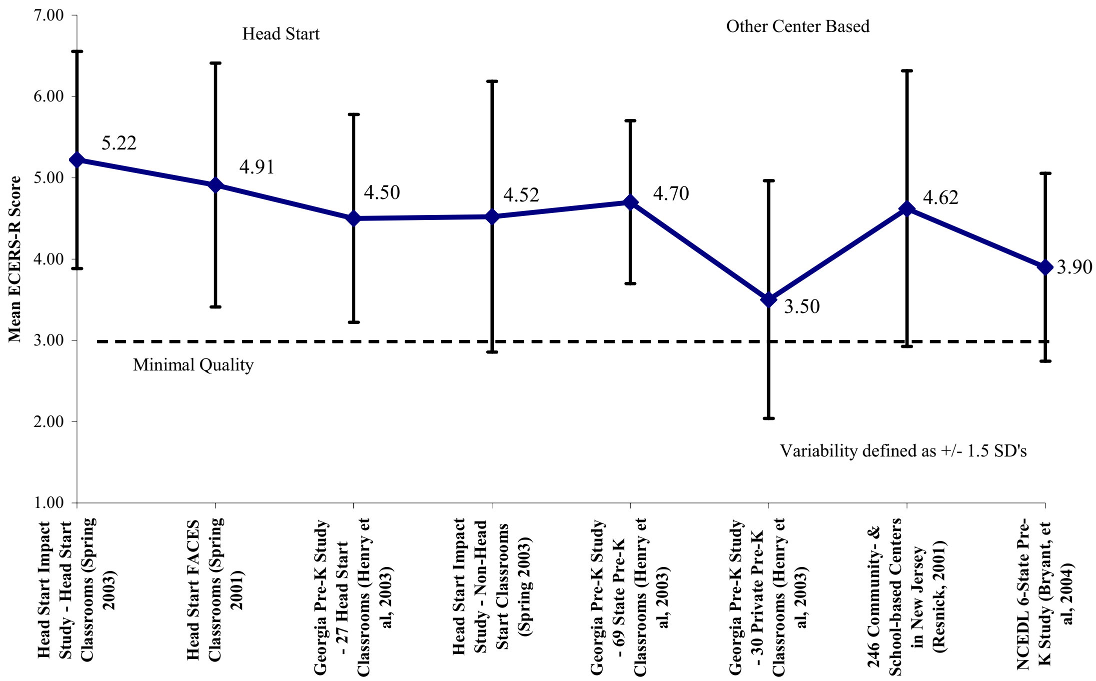
Exhibit 3.6: ECERS-R Scores for Children in 3- and 4-Year-Old Age Groups in Head Start and Other Center-Based Programs, Spring 2004 (Weighted Data)

Scale/Subscale	Head Start Centers Means [1]	Other Center-Based Programs Means[1]	Difference
3-Year-Old Group	**(N=1,154)**	**(N=187)**	
Total score	5.17	4.44	0.73***
Space & furnishings subscale	5.16	4.64	0.52***
Personal care routines subscale	5.45	4.65	0.80***
Language-reasoning subscale	5.11	4.44	0.67*
Activities subscale	4.67	3.66	1.01***
Interactions subscale	5.64	5.26	0.38
Program structure subscale	5.60	4.62	0.98***
4–Year-Old Group	**(N=860)**	**(N=269)**	
Total score	5.29	4.62	0.67***
Space & furnishings subscale	5.22	4.79	0.43**
Personal care routines subscale	5.58	4.76	0.82***
Language-reasoning subscale	5.26	4.67	0.59**
Activities subscale	4.70	3.96	0.74***
Interactions subscale	5.91	5.35	0.56**
Program structure subscale	5.84	4.82	1.02***

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.
[1] Center-level means calculated using individual child-level weights.

Exhibit 3.7 compares the Head Start Impact Study ECERS-R scores (for both Head Start and other center-based programs) with other studies using the overall ECERS-R score, to provide some context for the data from this study.[10] Studies of both Head Start and other center-based programs are provided. As shown, the ECERS-R total mean scores are similar, if not slightly higher, than reported scores in other studies of center-based care. Specifically, the total mean score for Head Start centers in FACES fall 2000 was 4.84; for the Georgia Pre-K Study of Head Start classes it was 4.5, as compared to 5.22 for the 3- and 4-year-old children in the Head Start Impact Study.

Exhibit 3.7: Measure of Classroom Quality in Head Start and Other Preschool & Child Care Settings.



Total ECERS-R mean scores for center-based programs (other than Head Start) range from 3.5-4.7, compared to 4.52 for the other center-based programs in the Impact Study.[11]

[10] To make these comparisons, the mean score combines the 3- and 4-year-old groups for Head Start centers and for other center-based programs.

[11] Henry, G.T., L.W. Henderson, D.P. Bentley, C.S. Gordon, A.J. Mashburn and D.K. Rickman. (2003). ***Report of the Findings From the Early Childhood Study: 2001-02.*** Atlanta, GA: Georgia State University, Andrew Young School of Policy Studies.

Bryant, D., O. Barbarin, R, Clifford, D, Early and R. Pianta. (2004). ***The National Center for Early Development and Learning: Multi-state Study of Pre-Kindergarten.*** Symposium Presentation at the Biennial Head Start Research Conference, Washington, DC.

Arnett Scale of Lead Teacher Behavior

A critical aspect of quality education for young children is a classroom environment in which children are nurtured, respected, and challenged.[12] The Arnett Scale of Lead Teacher Behavior is a 30-item scale that rates the lead teacher's behavior toward children in the class using a 4-point scale. A total score for the 30 items has been computed for the Head Start study, as well as for each of the five subscales measuring the lead teacher's behavior:

- **Sensitivity:** Ten items, a higher score indicates a teacher is more sensitive.
- **Harshness:** Nine items, a higher score indicates the teacher is **less** harsh.
- **Detachment:** Four items, a higher score indicates the teacher is **less** detached.
- **Permissiveness:** Three items, a higher score indicates the teacher is **less** permissive.
- **Independence:** Four items, a higher score indicates the teacher encourages the children to be independent and use self-help skills.

As shown in Exhibit 3.8, for children in the 4-year-old group in Head Start classrooms, teachers have significantly higher total Arnett scores than teachers in the other center-based classrooms (77.1 compared to 73.3). It is somewhat less certain, but still likely (p=.0504) that the overall score for the 3-year-old group in Head Start centers was higher than for the children in other center-based programs (75.09 compared to 70.30). Children in the 3-year-old group in Head Start had teachers who were rated more sensitive and who promoted more independence in children. For the 4-year-old group, Head Start teachers were rated as less harsh and who promoted more independence in children than non-Head Start teachers.

[12] Espinoza, L. (2004). "High-Quality Preschool: Why We Need It and What It Looks Like," ***Pre-School Policy Matters***, Issue 1, Nov. 2004, National Institute for Early Education Research.

Exhibit 3.8: Scores for the Arnett Scale of Lead Teacher Behavior, Head Start, and Other Center-Based Programs by Age Group, Spring 2003 (Weighted Data)

Scale/Subscale	Head Start Centers Mean[1]	Other Center-Based Programs Mean[1]	Difference
3-year-old Group	**(N=1,154)**	**(N=187)**	
Total score	75.09	70.20	4.89
Sensitivity subscale	23.18	20.57	2.53*
Harshness subscale	24.74	24.19	0.55
Detachment subscale	11.13	10.71	0.42
Permissiveness subscale	7.74	7.42	0.33
Independence subscale	8.78	7.37	1.41**
4-year-old Group	**(N=860)**	**(N=269)**	
Total score	77.10	73.30	3.79*
Sensitivity subscale	23.61	21.88	1.73
Harshness subscale	25.25	24.21	1.05*
Detachment subscale	11.39	11.08	0.31
Permissiveness subscale	7.95	7.68	0.27
Independence subscale	9.10	8.46	0.65*

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.
[1] Center-level means calculated using individual child-level weights.

Teacher Activities and Curriculum

Early education settings need directed and rich interactions between children and teachers in which teachers purposefully challenge and extend children's skills (Pianta, 2004)[13]. The higher scores on ECERS-R and Arnett for the Head Start classrooms are indicators that these programs are likely to promote better learning environments for children. To further explore the intentionality of classroom instruction, teachers in both Head Start and other center-based classrooms were asked how often they, or someone else, engaged the children in literacy, numeracy, or other activities such as arts and crafts, sports, indoor toys, or classroom chores. Teachers were asked to focus on how often these activities were done with the class in general and not to specifically focus on the children who were participating in the study. The respondents were given the choice of six responses: *Never*, *once a month or less, two or three times a month, once or twice a week, three or four times a week, every day*. These responses were collapsed into three broader categories: *Never*, *Sometimes* (included once a month or less, two or three

[13] Pianta, R. (2004). "Transitioning to School: Policy, Practice, and Reality." ***The Evaluation Exchange***, Vol. 10, #2, 2004, Harvard Family Research Project.

times a month, and once or twice a week) and *Frequently* (three or four times a week and every day). Many of the activities selected were intended to focus on children's active involvement in learning. The results are discussed below.

Language and Literacy Activities

Early literacy may be promoted through learning letters, phonics, and exposure to a rich vocabulary and variety of printed materials. Preschool children need to learn certain concepts to become competent readers, including vocabulary and language; phonological awareness; knowledge of print, letters and words; comprehension; understanding books; and literacy for enjoyment (Snow, Burns, & Griffin, 1998).[14] Teachers were asked about the frequency with which they implemented 11 different activities aimed at these concepts, including naming and writing letters, learning letter sounds and rhyming words, and understanding story and print concepts. The 3-year-old group children in Head Start classrooms were provided significantly more instruction for 5 of these 11 activities (see Exhibit 3.9). Specifically, teachers reported implementing activities in discussing new words, oral comprehension, and writing skills. However, for the teachers of children in the 4-year-old group, no significant differences were found between Head Start and other center-based classrooms.

Math Activities

The National Standards in Mathematics identify key components of math instruction for preschool children, including number concepts, patterns and relationships, shapes and spatial sense, and measurement. Teachers were asked about a number of these concepts as shown in Exhibit 3.10. Children in both age groups in Head Start classrooms were more frequently provided math activities than children in other center-based classrooms. As shown in Exhibit 3.10, Head Start teachers of the 3-year-old group more frequently used almost all of the eight activities (6 of the 8 showed significant differences). Head Start teachers of the 4-year-old group were significantly more likely to use math games, music, and dance to learn math concepts and activities emphasizing measurement. Significant differences were not found in the other areas among teachers of the 4-year-old group.

[14] Snow, C.E., M.S. Burns, and P. Griffin (Eds). (1998). ***Preventing Reading Difficulties in Young Children.*** Washington, DC: National Academy Press.

Other Types of Activities

Teachers were also asked how frequently they work on art and craft activities, play with games or toys indoors, play sports or exercise, and have children help with chores to promote independence. Children in both age groups in Head Start classrooms were more frequently provided art and craft activities and more frequently engaged in chores to promote independent behavior (see Exhibit 3.11). Indoor games and sports were very frequently used in both Head Start and non-Head Start centers.

Use of Curriculum

Curriculum plays an important role in shaping how the classroom day is structured and the types of activities the teachers focus on in class. For children in other non-Head Start center-based program classrooms, approximately 14 percent of the children in the 3-year-old group and 17 percent of children in the 4-year-old group were in classrooms that **did not use** a curriculum (see Exhibit 3.12). This compares to about 2 percent of the children in the 3-year-old group and 4 percent of the children in the 4-year-old group that were in Head Start classrooms.

The philosophies and scope of curricula commonly used in preschool classrooms often vary widely. As shown in Exhibit 3.12, a high percentage of children in Head Start classrooms are exposed to common curricula. More than three-quarters of the 3-year-old group and approximately 80 percent of children in the 4-year-old group were in classrooms that used either High Scope or The Creative Curriculum. In FACES 2000, similar findings indicated that the majority of teachers in Head Start used either The Creative Curriculum or High Scope.[15] Thus, there has been consistency in the use of these two curricula over time in the Head Start program. Both of these curricula have similar philosophies that support developmentally appropriate practices that encourage children to make choices about materials and activities during the day. This philosophy encourages children to actively learn concepts by playing with or manipulating materials (Dodge, Colker, & Heroman, 2002)[16]. This matches the earlier finding that children in Head Start center classrooms were exposed to developmentally appropriate "hands on" activities more frequently.

---

[15] Administration on Children & Families (2003). Retrieved from: http://www.acf.hhs.gov/programs/opre/hs/faces/report/faces00_4thprogress/faces00_title.html.

[16] Dodge, D.T., L.J. Colker & C. Herman. (2002). ***The Creative Curriculum for Pre-School.*** Washington, DC: Teaching Strategies, Inc.

Exhibit 3.9: Percentage of Children in Head Start and Other Center-Based Classrooms by Frequency of Use of Language and Literacy Activities, 3- and 4-Year-Old Age Groups, Spring 2004 (Weighted Data)

Language and Literacy Activities	3-Year-Old Group						4-Year-Old Group					
	Head Start Centers			Other Center-Based Programs			Head Start Centers			Other Center-Based Programs		
	Never	Some-times	Frequently	Never	Some-times	Frequently	Never	Sometimes	Frequently	Never	Sometimes	Frequently
Name Letters	1.1	20.7	78.2	9.9	24.5	65.6	0.9	15.7	83.4	0.8	20.3	78.9
Write letters (3-year-old group)**	4.0	39.3	56.6	21.8	34.9	43.3	2.3	30.9	66.7	5.2	43.2	51.7
Letter sounds (phonics)	3.5	33.9	62.6	11.2	30.5	58.2	3.1	27.6	69.3	2.9	32.2	64.8
Write/spell name (3-year-old group*)	3.3	27.4	69.3	18.4	25.8	55.9	1.7	20.0	78.3	2.2	34.4	63.4
Discuss new words (3-year-old group*)	0.5	23.9	75.5	7.7	41.3	51.0	0.6	22.3	77.1	0.6	33.9	65.5
Have children tell stories	0.5	42.9	56.5	4.3	55.2	40.5	0.0	48.7	51.3	2.1	57.5	40.4
Read to children (show print)	0.4	15.0	84.6	3.9	23.3	72.8	0.3	16.5	83.2	0.2	19.6	80.2
Retell/make up stories (3-year-old group*)	0.9	48.8	50.4	8.4	59.9	31.6	0.1	53.6	46.4	4.1	61.4	34.6
Show how to read a book	0.7	28.5	70.7	5.0	34.9	60.0	0.3	24.2	75.5	2.2	29.6	68.1
Teach directional words	0.4	33.3	66.3	4.9	39.2	55.9	0.0	38.5	61.5	0.7	45.4	53.8
Learn rhyming words	5.5	48.7	45.8	15.1	55.3	29.5	2.1	47.8	50.1	3.8	56.7	37.5

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit 3.10: Percentage of Children in Head Start and Other Center-Based Classrooms Where Math Activities Are Used by Frequency of Activity, 3- and 4-Year-Old Groups, Spring 2004 (Weighted Data)

Math Activities	3-Year-Old Group						4-Year-Old Group					
	Head Start Centers			Other Center-Based Programs			Head Start Centers			Other Center-Based Programs		
	Never	Sometimes	Frequently	Never	Sometimes	Frequently	Never	Sometimes	Frequently	Never	Sometimes	Frequently
Count aloud	0.0	4.5	95.5	2.0	11.8	86.2	0.0	4.9	95.1	0.0	6.5	93.5
Calendar/days of the week	2.8	11.3	85.8	5.2	21.4	73.4	1.8	15.3	82.9	0.5	15.7	83.9
Work with shape blocks (3-year-old group)**	0.4	11.6	88.0	1.8	23.4	74.9	0.1	16.9	82.9	0.5	18.4	80.8
Count small toys (3-year-old group)**	0.0	14.3	85.8	5.1	26.5	68.5	0.0	16.1	83.9	0.3	25.6	74.2
Play math games (3-year-old group, 4-year-old group*)**	0.7	34.7	64.7	8.6	47.9	43.6	0.4	34.4	65.2	4.5	44.5	50.1
Use music to learn math (3-year-old group, 4-year-old group*)**	3.8	36.8	59.4	21.8	42.9	35.3	2.2	47.1	50.7	20.1	42.5	37.4
Use dance to learn math (3-year-old group*, 4-year-old group*)	3.5	43.9	52.5	16.7	51.8	31.6	2.0	46.5	51.5	16.1	51.8	32.1
Use rulers/ measuring cups (3-year-old group*, 4-year-old group***)**	3.7	46.5	49.9	16.0	54.2	29.8	0.1	53.7	46.3	8.9	69.9	22.1

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit 3.11: Percentage of Children in Head Start and Other Center-Based Classrooms Where Other Activities Are Used by Frequency of Activity, 3- and 4-Year-Old Groups, Spring 2004 (Weighted Data)

Other Activities	3-Year-Old Group						4-Year-Old Group					
	Head Start Centers			Other Center-Based Programs			Head Start Centers			Other Center-Based Programs		
	Never	Sometimes	Frequently	Never	Sometimes	Frequently	Never	Sometimes	Frequently	Never	Sometimes	Frequently
Work on arts and crafts (3-year-old group*, 4-year-old group*)	0.2	9.3	90.6	1.8	22.5	75.8	0.5	9.7	89.9	0.0	22.6	77.4
Play indoor games or toys	0.0	5.3	94.7	1.0	8.0	90.9	0.0	1.7	98.3	0.0	3.3	96.7
Play sports or exercise	0.0	6.3	93.7	2.8	13.6	83.7	0.0	5.9	94.1	0.2	13.2	86.6
Help with chores (3-year-old group, 4-year-old group*)**	0.1	5.1	94.7	3.9	19.7	76.3	0.1	3.9	95.9	2.3	10.5	87.2

* = $p \le 0.05$, ** = $p \le 0.01$, *** = $p \le 0.001$.

There was significantly more variation in curricula among the other center-based teachers surveyed for both age groups (see Exhibit 3.12). Only 41 percent of the 3-year-old group was in classrooms in which teachers reported using High Scope or The Creative Curriculum. Similarly, for children in the 4-year-old group, approximately 46 percent of the children were in classrooms that used High Scope or The Creative Curriculum. A wide array of other curricula were mentioned by the teachers in the other center-based classrooms, with state-developed curricula mentioned most frequently.

Exhibit 3.12: Percentage of Children in Head Start and Other Center-Based Programs by Type of Curriculum Used in the Classroom, 3- and 4-Year-Old Groups, Spring 2004 (Weighted Data)

Type of Curriculum	3-Year-Old Group (N = 1,259)		4-Year-Old Group (N = 1,067)	
	Head Start Centers	Other Center-Based Programs	Head Start Centers	Other Center-Based Programs
High Scope or Creative Curriculum	76.7***	41.0	79.7**	46.2
Other curriculum	21.6	45.1	16.8	36.5
No curriculum	1.7	13.9	3.5	17.3
Total percent	100%	100%	100%	100%

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Chapter 4: Overview of Methods for Analyzing Impacts on Children and Families

Purpose of Interim Analysis

This chapter describes key aspects of the analysis methods used in estimating Head Start's impact on children's cognitive and social-emotional development, health, and parenting practices. The narrative begins with a discussion of the key outcome domains, constructs, and measures that were selected for this examination of program impacts including procedures used to create scales and other derived variables. It also reviews the methods used to estimate overall average program impacts, and impacts for particular subgroups and concludes with a discussion of ways to deal with the nonadherence to random assignment discussed in Chapter 2.

Outcome Domains and Measures

As discussed in Chapter 1, a wide variety of data sources and measures are being used to assess the impact of Head Start on fostering and enhancing child development, including direct child assessments, parent/primary caregiver interviews, interviews with providers of early care services used by participating study children, and observations of children's early care settings. For this report, a selected set of key outcome measures in the cognitive, socio-emotional, health, and parenting domains were used for this initial assessment of the impact of Head Start. These are described below and summarized in Exhibit 4.1[1] for the combined treatment and comparison groups presented separately by age group (3- and 4-year-olds), based on information collected in spring 2003:

1. Cognitive Domain:

- **Pre-reading skills.** These skills focus primarily on letter recognition, an important step toward becoming a proficient reader**.** This domain is measured by the Woodcock-Johnson-III Letter-Word Identification subtest and the Letter Naming Task.
- **Pre-writing skills.** Children's ability to draw shapes and write letters and words is assessed**.** This domain is measured by the Woodcock-Johnson III Spelling subtest and McCarthy Draw-a-Design Test.
- **Vocabulary knowledge.** This skill is indicative of children's oral language development and general knowledge. This domain is measured by the PPVT-III and the Color Naming Task.

---

[1] For certain cognitive measures, information is provided for both Item Response Theory maximum likelihood values and standard scale scores.

Exhibit 4.1: Spring 2003 Outcome Measures, Data Source or Scoring Method, and Descriptive Statistics for the Combined Sample [1]

Outcome Measure	Source/Scoring	Mean[2] Standard Deviation Range	
		3-Year-Old Group	4-Year-Old Group
Cognitive Domain			
Peabody Picture Vocabulary Test-III (adapted)	Child Assessment; IRT scoring of the adapted version	M: 252.01 (82.17) SD: 35.56 R: 148-382	M: 292.63 (87.55) SD: 38.75 R: 174-414
Comprehensive Test of Phonological and Print Processing (CTOPPP): Elision	Child Assessment; IRT scoring	M: 241.59 SD: 43.63 R: 131-379	M: 274.47 SD: 48.14 R: 132-385
Letter Naming Task	Child Assessment; Number of letters identified correctly	M: 4.71 SD: 7.20 R: 0-26	M: 10.40 SD: 9.70 R: 0-26
Color Naming/Identification	Child Assessment; Number of colors identified correctly	M: 13.49 SD: 6.90 R: 0-20	M:16.78 SD: 5.30 R: 0-20
Counting Bears	Child Assessment; Measure of one-to-one counting	M: 2.77 SD: 1.32 R: 1-5	M: 3.68 SD: 1.34 R: 1-5
McCarthy Scales of Children's Abilities: Draw-a-Design	Child Assessment; Number of shapes drawn correctly	M: 3.13 SD: 1.17 R: 0-12	M: 4.46 SD: 2.03 R: 0-15
Woodcock-Johnson III Tests of Achievement: Letter-Word Identification	Child Assessment; W score generated by the Woodcock-Johnson Compuscore and Profiles Program	M: 303.77 (91.53) SD: 25.05 R: 264-392	M: 322.41 (92.44) SD: 27.80 R: 264-408
Woodcock-Johnson III Tests of Achievement: Spelling	Child Assessment; W score generated by the Woodcock-Johnson Compuscore and Profiles Program	M: 345.11 (92.47) SD: 22.58 R: 277-426	M: 369.66 (90.99) SD: 25.44 R: 277-442
Woodcock-Johnson III Tests of Achievement: Applied Problems	Child Assessment; W score generated by the Woodcock-Johnson Compuscore and Profiles Program	M: 375.43 (88.16) SD: 28.39 R: 318-436	M: 395.98 (87.57) SD: 25.53 R: 318-436
Woodcock-Johnson III Tests of Achievement: Oral Comprehension	Child Assessment; W score generated by the Woodcock-Johnson Compuscore and Profiles Program	M: 435.48 (92.25) SD: 14.04 R: 418-489	M: 443.52 (90.68) SD: 17.92 R: 418-489
Test de Vocabulario en Imágenes Peabody (adapted)[3]	Child Assessment; IRT scoring of the adapted version	M: 250.18 (90.31) SD: 40.41 R: 160-383	M: 293.56 (88.18) SD: 43.80 R: 149-442
Batería Woodcock-Muñoz Pruebas de aprovechamiento-Revisada: Identificación de letras y palabras[3]	Child Assessment; W score calculated from the Woodcock-Munoz scoring table in the Test Record	M: 351.17 (93.49) SD: 12.52 R: 316-392	M: 357.54 (86.21) SD: 11.50 R: 316-423
Parent (reported) Emergent Literacy Scale	Parent Interview; Sum of five items	M: 2.61 SD: 1.45 R: 0-5	M: 3.55 SD: 1.39 R: 0-5

Exhibit 4.1: (continued)

Outcome	Source/Scoring	Mean Standard Deviation Range	
		3-Year-Old Group	4-Year-Old Group
Socio-emotional Domain			
Social Skills and Positive Approaches to Learning	Parent Interview; Sum of seven items	M: 12.39 SD: 1.72 R: 4-14	M: 12.48 SD: 1.72 R: 4-14
Total Child Behavior Problems Scale	Parent Interview; Sum of 12 items	M: 6.01 SD: 3.66 R: 0-22	M: 5.70 SD: 3.59 R: 0-19
Aggressive Behavior Scale	Parent Interview; Sum of four items	M: 3.01 SD: 1.72 R: 0-8	M: 2.79 SD: 1.69 R: 0-8
Hyperactive Behavior Scale	Parent Interview; Sum of three items	M: 1.85 SD: 1.55 R: 0-6	M: 1.74 SD: 1.47 R: 0-6
Withdrawn Behavior Scale	Parent Interview; Sum of three items	M: 0.57 SD: 0.95 R: 0-6	M: 0.68 SD: 0.96 R: 0-6
Social Competencies Checklist	Parent Interview; Home inventory from the Developing Skills Checklist; Sum of 12 items	M: 10.98 SD: 1.32 R: 0-12	M: 11.04 SD: 1.32 R: 1-12
Parenting Practices Domain			
Parent used time out in the last week	Parent Interview; One item	M: 0.64 SD: 0.48 R: 0-1	M: 0.64 SD: 0.48 R: 0-1
Number of times parent used time out in the last week	Parent Interview; One item	M: 1.77 SD: 2.25 R: 0-28	M: 1.68 SD: 2.50 R: 0-100
Parent spanked child in the last week	Parent Interview; One item	M: 0.45 SD: 0.50 R: 0-1	M: 0.37 SD: 0.48 R: 0-1
Number of times parent spanked child in the last week	Parent Interview; One item	M: 0.90 SD: 1.52 R: 0-21	M 0.70 SD: 1.21 R: 0-20
Parental Safety Practices Scale	Parent Interview; Average score for five items	M: 3.71 SD: 0.33 R: 2-4	M: 3.72 SD: 0.33 R: 2-4
Removing Harmful Objects Scale	Parent Interview; Average score for seven items	M: 3.89 SD: 0.32 R: 1-4	M: 3.89 SD: 0.32 R: 1-4
Restricting Child Movement Scale	Parent Interview; Average score for four items	M: 3.89 SD: 0.29 R: 1-4	M: 3.88 SD: 0.31 R: 1-4
Safety Devices Scale	Parent Interview; Average score for two items	M: 3.34 SD: 0.75 R: 1-4	M: 3.39 SD: 0.77 R: 1-4
Family Cultural Enrichment Scale	Parent Interview; Sum of seven items	M: 3.65 SD: 1.41 R: 0-7	M: 3.95 SD: 1.43 R: 0-7
How many times child was read to in the last week by parent or other family member	Parent Interview; One item	M: 2.86 SD: 0.94 R: 1-4	M: 2.88 SD: 0.95 R: 1-4

Exhibit 4.1: (continued)

Outcome	Source/Scoring	Mean Standard Deviation Range	
		3-Year-Old Group	**4-Year-Old Group**
Health Domain			
Child seen by dentist since last September	Parent Interview; One item	M: 0.60 SD: 0.49 R: 0-1	M: 0.65 SD: 0.48 R: 0-1
Overall child's health status	Parent Interview; One item	M: 0.78 SD: 0.41 R: 0-1	M: 0.80 SD: 0.40 R: 0-1
Child had injury in last month requiring medical treatment	Parent Interview; One item	M: 0.09 SD: 0.28 R: 0-1	M: 0.12 SD: 0.32 R: 0-1
Child has health insurance	Parent Interview; One item	M: 0.92 SD: 0.27 R: 0-1	M: 0.88 SD: 0.32 R: 0-1
Child has place for routine medical care	Parent Interview; One item	M: 0.98 SD: 0.14 R: 0-1	M: 0.97 SD: 0.17 R: 0-1
Child has condition that requires ongoing medical care	Parent Interview; One item	M: 0.13 SD: 0.34 R: 0-1	M: 0.11 SD: 0.32 R: 0-1
Child has an unmet health care need	Parent Interview; One item	M: 0.02 SD: 0.13 R: 0-1	M: 0.03 SD: 0.18 R: 0-1

[1] The combined sample includes children assessed in all languages in fall 2002 (including English, Spanish, and other languages) and in English in spring 2003.

[2] Scores in parentheses are standard scores for tests where available.

[3] Indicates administered to Spanish-speaking children only in the combined sample.

- **Oral comprehension and phonological awareness.** This includes the child's ability to understand and make inferences from phrases and sentences spoken in English and to understand that spoken sentences are made of component words, compound words are made up of simpler words, and that words are made up of component syllables and sounds (phonemes). This domain is measured by the Woodcock-Johnson III Oral Comprehension subtest and the Elision subtest of the Comprehensive Test of Print and Phonological Processing─Preschool Edition (CTOPPP).
- **Early math skills.** Child assessments include basic math skills and understandings that are essential for the development of more advanced quantitative capabilities. This domain is measured by the Woodcock-Johnson III Applied Problems subtest and the Counting Bears Task.
- And, finally, a measure of **parent's perceptions** of their child's literacy skills, using information from the parent interview.

2. Social-Emotional Domain:

- **Social skills and approaches to learning.** Parents were asked to rate their child's social skills and positive approaches to learning.[2] Social skills focused on cooperative and empathic behavior, such as, "Makes friends easily," "Comforts or helps others," and "Accepts friends' ideas in sharing and playing." Approaches to learning deal with curiosity, imagination, openness to new tasks and challenges, and having a positive attitude about gaining new knowledge and skills. Examples include, "Enjoys learning," "Likes to try new things," and "Shows imagination in work and play." The two scales are based on an instrument used in the Head Start Family and Child Experiences Survey (FACES).[3]
- **Social competencies.** Parents were asked to provide information on social capabilities using a Social Competencies Checklist, also used in FACES 2000. The checklist consisted of 12 items; for each item, the parent was asked to report whether the child engaged in that behavior or exhibited that attribute "regularly" or "very rarely or not at all." Examples of the items included, "Shares newly learned ideas," "Takes care of personal belongings," "Helps with simple household tasks," and "Notices when others are happy, sad, angry." The total scale score could range from zero (all items rated "rarely or not at all") to 12 (all items rated "does regularly").
- **Problem behavior.** Parents were asked to rate their children on items dealing with aggressive or defiant behavior such as, "Hits and fights with others," "Has temper tantrums or hot temper," and "Is disobedient at home." Other items dealt with inattentive or hyperactive behavior, including, "Can't concentrate, can't pay

---

[2] For each item, the parent was asked to judge whether the behavioral description was "not true," "sometimes true," or "very true" of the child. There were seven items in this scale, and scores could range from zero (meaning all the items were rated "not true" of the child) to 14 (meaning all the items were rated "very true" of the child). Mean scores on the scale obtained from parents of Head Start children in the Head Start Impact Study were closely comparable to mean scores obtained from parents of an independent national sample of Head Start children in FACES 2000.[2] As in FACES, social skills and positive approaches to learning scores tended to be skewed toward the higher end of the range because parents tended to rate their children as exhibiting most of the positive attributes asked about in the rating instrument. Nonetheless, the scale has shown significant relationships with other measures of children's social development and with relevant child and family characteristics.

[3] Administration Children and Families. (2001). Retrieved 10/15/04 from: http://www.acf.hhs.gov/programs/core/ongoing_research/faces/faces_instruments.html/.

attention for long," and "Is very restless and fidgets a lot." A third set of items dealt with shy, withdrawn, or depressed behavior, e.g., "Feels worthless or inferior," and "Is unhappy, sad, or depressed." For each item, the parent was asked to judge whether the behavioral description was "not true," "sometimes true," or "very true" of the child. The **Total Behavior Problem** scale derived from parent ratings contained 14 rating items, and the total scale score could range from zero (all items marked "not true") to 28 (all items marked "very true"). The **Aggressive Behavior** subscale contained four items and could range from zero to eight. The **Hyperactive Behavior** subscale contained three items, and scores could range from zero to six. The **Withdrawn Behavior** subscale contained three items, and scores could range from zero to six. These scales were also used in FACES 2000, and their development was based on prior work by Rutter, Achenbach, Zill and Peterson, and others (see ACF, 2001). The mean scores obtained in the Head Start Impact Study were very comparable to mean scores obtained from parents of an independent national sample of Head Start children surveyed for FACES 2000.[4]

3. Health Domain:

- **Access to health care.** Parents were asked to report on various health care services, two of which are used in this report:
  - **Whether the child has health insurance.** Parents were asked if the child was covered by Medicaid or a state health insurance program, or by health insurance through their job or the job of another employed adult.
  - **Whether the child has received dental care.** Parents were asked if the child had ever seen a dentist.
- **Child's health status.** Parents were asked to report on their child's health status:
  - **Child's health status (excellent or very good).** Parents were asked if, overall, the child's health was excellent, very good, good, fair, or poor. This outcome was coded "yes" for those who reported that their child's health was excellent or very good.
  - **Whether the child needs ongoing medical care.** Parents were asked if their child had an illness or condition that requires regular ongoing medical care.
  - **Whether child received medical care for an injury in the last month.** Parents were asked how many times their child, in the last month, had seen a doctor or other medical professional or visited a clinic or emergency room for an injury. This outcome was coded yes if the parent reported any such occurrences in the last month.

4. Parenting Practices Domain:

- **Educational activities.** Parents were asked to report on the types of educational activities they did with their child:

[4] Zill, N., et al. (2003). ***Head Start FACES 2000: A Whole-Child Perspective on Program Performance. Fourth Progress Report.*** Washington, DC: Administration Children and Families, US Department of Health and Human Services.

- **Reading to the child at home.** Parents reported on the item "How many times have you or someone in your family read to [CHILD] in the past week?" Possible responses range from 1 (not at all) to 4 (every day).

- **Cultural enrichment activities.** Parents reported on a 7-item checklist of activities the parent, or another family member, may have done with the child during the past month. The seven activities include going to a movie; play or concert; art gallery or museum; playground, park, or zoo; community, ethnic, or religious event; and talking about family or cultural heritage and going on errands. A total score was computed by summing the number of different activities the parent and child participated in together, with a possible score of 0 (none) to 7 (all).

- **Discipline strategies.** Parents reported on the following:
  - **Use of physical discipline.** 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?" For parents who responded yes, the **Frequency of physical discipline** was also created from parent reports on the item "About how many times in the past week?" Responses ranged from 0 to 21 times.
  - **Use of time out.** Parents reported on the item "Have you used 'time out' or sent [CHILD] to his/her room in the past week for not minding?" For parents who responded yes, the **Frequency of time out** was also created from parent reports on the item "About how many times in the past week?" Responses ranged from 0 to 100 times.

- **Child safety practices.** Parents reported on a 10-item scale that assessed how often the 10 different safety precautions were used, including keeping harmful objects out of reach, using car seats, supervising the child during bath time, and having a first aid kit and working smoke detector at home. Possible responses ranged from 1 (never) to 4 (always). In addition to a total overall scale score, exploratory factor analyses yielded three separate subscales that were also used in the analysis: removing harmful objects from the home, restricting child movements from dangerous situations, and having safety devices available for the child.

Creation of Test Scores and Scales

As noted in Exhibit 4.1, IRT analysis was used to develop the adapted (shortened) versions of the PPVT-III and TVIP to significantly reduce the time required to test individual children (i.e., reducing the burden on the child). IRT analysis achieves this goal by treating test items as interchangeable components that can be added or substituted without altering the underlying test scale, i.e., higher ability children did not have to be administered easier items, and lower ability children did not have to be administered more difficult items to get a reliable test score for each child.

IRT analysis was also used to score child assessments for the PPVT-III, TVIP, and CTOPPP Elision tests. The advantage of IRT analysis for scoring is that it uses the actual pattern of right, wrong, and omitted responses to the items administered in an assessment and uses the item difficulty, discrimination, and guessing behavior to place each child on a continuous ability scale. In this case, data from the multiple waves of existing FACES data collection were used to conduct the IRT analysis. If an assessment is shortened (as in the case of the PPVT-III or TVIP). IRT analysis also has advantages over the use of simple raw scores. By using the overall pattern of right and wrong answers, and the characteristics of each item to estimate ability, IRT analysis can compensate for the possibility that a low-ability child will correctly respond to several difficult items by guessing. Unlike raw scores, which treat omitted items as if they had been answered incorrectly, IRT procedures use the pattern of actual responses to estimate the probability of correct responses for all assessment questions, including any omitted items. The Compuscore and Profiles Program (Riverside Publishing, 2001) was used to score the Woodcock-Johnson III subtests, and publisher look-up tables (Riverside Publishing, 1990) were used to score the Woodcock-Muñoz. The total number of correct responses (or raw score) was used for the Letter Naming Task, Color Naming/Identification, Counting Bears, and the McCarthy Draw-a-Design subtest.

In addition to the direct child assessments discussed above, the following additional scales were developed using items from the parent interview:[5]

- **Problem behaviors.** As discussed above, scales were developed to assess children's social-emotional development. The items making up these ratings were drawn from three measures of children's positive behavior and behavior problems: the Entwisle Scale of Personal Maturity (Entwisle, Alexander, Cadigan, & Pallis, 1987), the Child Behavior Checklist for Preschool-Aged Children (Achenbach, Edelbrock, & Howell, 1987), and the Home Inventory in the Developing Skills Checklist (CTB/McGraw-Hill, 1990).
- **Maternal depression and locus of control.** These two measures were used as covariates in the statistical models and, in the case of depression, as a moderator of program impact (see following discussion). The Depression Scale is derived from the CES-D Depression Scale (Ross, Mirowsky, & Huber, 1983), and the Locus of Control Scale is derived from the Pearlin Mastery Scale (Pearlin & Schooler, 1978).
- **The Parent Emergent Literacy Scale (PELS).** PELS is a parent-report on five literacy items originally developed for use in FACES 2000: child can recognize most/all of the letters of the alphabet; child can count to 20; child pretended to write his/her name in the last month; child can write his/her first name; and child can identify the primary colors.

[5] Citations for these scales are in Appendix 4-1.

IRT was used to generate the scale scores for the caregiver's depression and locus of control scales. For the remaining scales, the scale score results from a summation of the item responses for each scale.

The Analysis Sample

The sample used in this report to estimate impacts on children and families was chosen to maximize the data available by including every completed child assessment and parent interview from the spring 2003 wave of data collection (the end of the first Head Start program year). Observations were compiled independently for child assessments and parent interviews, and information was included from one of these sources even when the other source was missing. For this reason, and also due to item nonresponse for specific questions in completed questionnaires, sample sizes are not identical for all analyses, i.e., different outcome variables involve slightly differing numbers of observations. The comparability of the Head Start and non-Head Start samples established at random assignment is maintained to the greatest extent possible in each instance by adjusting the initial sampling weights to offset observable differences between respondents and nonrespondents at baseline (see the discussion below and Appendix 1.2 for details).

Rather than dropping observations with missing data on background variables from the fall 2002 data collection, a statistical "hot-deck" procedure was used to impute missing background variables for cases with either (1) no fall 2002 parent interviews, or (2) incomplete fall 2002 data caused by item nonresponse. Appendix 4.1 provides details of the imputation process, including initial missing data rates for all the imputed variables.

The set of completed questionnaires and assessments was divided into two separate samples, one for children entering Head Start 1 year before anticipated kindergarten entry—referred to as the 4-year-old group—and one for children entering Head Start 2 years prior to expected kindergarten entry—the 3-year-old group. This corresponds to the structure of the original random assignment, which was done separately for the two age groups to allow a separate experimental examination of each group of newly entering children. Analysis weights were established separately for the child assessments and for parent interviews. The analysis weights (which initially were based on the probability of selection into the study sample during random assignment) were adjusted to compensate for nonrespondents by increasing the weight for responding children with similar individual and family background characteristics on variables measured for **all** randomly assigned cases in 2002, prior to random assignment. (See Appendix 1.2 for details of the weight-adjustment procedures.)

The weighted data, therefore, intend to represent the same universe for all spring 2003 outcomes examined: the national population of newly entering 3-year-olds and, separately, the national population of newly entering 4-year-olds. For some purposes, the universe at each age level is divided by the primary language used to assess the child in fall 2002 and spring 2003. As discussed in Chapter 1, children who could not complete all the assessment batteries in English had their assessments primarily administered in Spanish or, for a small fraction of the sample, some other non-English language. In examining Head Start's impact on child cognitive and social-emotional development, we report separately on the set of children assessed initially in Spanish and the children assessed initially in English or some other language.[6] Like all subgroups defined by characteristics independent of the intervention and not affected by random assignment, these subsamples are, but for chance,[7] well-matched between children originally randomized into the Head Start versus non-Head Start groups and represent a valid experimental examination in and of themselves. Thus, the separate language-of-assessment analyses provide equally unbiased measures of Head Start's impact on that particular subpopulation as does the study as a whole for the full population.

In general, all the analyses described in this chapter encompass children assessed in all languages in fall 2002 (including Spanish, English, and other), other than those in Puerto Rico, and are carried out separately for the 3-year-old and 4-year-old cohorts. The one exception concerns spring outcome measures collected for children assessed initially (i.e., in fall 2002) in Spanish. As noted elsewhere, child assessments used in spring 2003 for these children were supplemented with two cognitive assessments designed specifically for Spanish-speaking children and administered **only to those children whose original fall assessments were conducted in Spanish,** i.e., the TVIP (adapted) and the Woodcock-Muñoz Letter-Word Identification Test. The current report includes these two tests in the impact analyses for those children whose initial fall assessments were conducted primarily in Spanish.

---

[6] Both sets of children had advanced sufficiently in their English language skills by spring 2003 to be administered follow-up assessments primarily in English (with continued use of several measures administered in Spanish) for use as outcome data for the impact analysis, with the exception of children in Puerto Rico. All Puerto Rican children in the sample spoke Spanish as their native language at the time of random assignment and continued to be assessed in Spanish in the spring. For this reason, Puerto Rico sample members, and hence the Puerto Rico-based portion of the national Head Start program, are not included in the report: The cognitive measures thought crucial to gauging Head Start's impact are not available for these children on a comparable basis at this early age. Puerto Rico will be added to the analysis sample in subsequent years.

[7] In addition to chance, the comparability of the Head Start and non-Head Start samples for the different language groups will depend on the success of the nonresponse weight adjustments made to the overall sample to deal with possible differential nonresponse in the spring 2003 data collection.

Methods of Estimating Head Start Impacts

The impact of Head Start is assessed from three points of view, reflecting three philosophies for arriving at the best evidence from a randomized experimental design: (1) differences in average outcomes, (2) differences in outcomes adjusted for fall 2002 demographic characteristics, and (3) differences in outcomes adjusted for both children's demographic characteristics and fall 2002 "starting points" on the outcome measures used in the particular analysis (e.g., fall PPVT measure). Each method is discussed below.

Difference in Average Outcomes

The National Head Start Impact Study, like other evaluations that use random assignment to allocate slots to program participants, provides a framework for attributing child outcomes to the effects of the program, rather than to other factors that may influence child development. Unlike pre-test/post-test analyses and other comparison group approaches, this framework makes accurate impact measurement possible without considering any individual child's starting point. If enough individuals are randomized to the Head Start and non-Head Start groups, and if all randomized individuals are included in the follow-up analysis, important differences in later outcomes are almost certain to result from the intervention being examined rather than other factors. In rare instances these groups can differ by chance alone on background factors affecting outcomes. However, statistical tests are used to decide if outcome differences are significant, thereby reducing the probability of reaching false conclusions to 5 percent or less. Actual measurement, and adjustment for possible chance differences in starting points, is not essential under this design (although it can be useful for certain reasons, as discussed below).

The simplicity of the basic Head Start/non-Head Start comparison of spring outcomes, without recourse to other data, provides a powerful motivation for evaluating program impacts in just this way. The transparency of the methodology, and its lack of dependence on sometimes complex statistical methods, makes these "difference-in-means" results good candidates as initial measures of Head Start's impact. Appendix 4.2 presents the most basic version of this analysis, contrasting the average outcome level for the Head Start group with the average outcome level for the non-Head Start group using unweighted data. This is as close to a simple "randomize and see what happens" approach to experimental evaluation as possible. However, the unweighted estimates can be biased because they do not take into account the differential probabilities of selection of children in the sample. The child weights account for the sampling of PSUs, grantee/delegate agencies, centers, and children within centers so that the study sample can be used to represent the national Head Start population.

These weighted difference-in-means impact estimates are reported in Chapters 5-8 and are also included in Appendix 4.2 for comparison purposes. Statistical tests determine which of the measured outcome differences between Head Start and non-Head Start children can be considered real impacts rather than simply due to sampling error. For continuous outcome variables (e.g., PPVT III scale score), the tests are based on ordinary least-squares (OLS) regression models that replicate the difference-in-means calculation by expressing spring 2003 outcomes as the sum of an intercept term and a shift in the intercept produced by a dummy variable for inclusion in the Head Start group.[8] For discrete outcome variables (e.g., use of dental care), logistic regressions were used to do equivalent computations converted to a scale of 0 to 1 to obtain measures of impact on the probability that a particular outcome occurs (e.g., getting a dental check-up).[9] Appendix 4.3 describes both procedures in detail, including a formal statement of the regression equation in mathematical notation.

Outcomes Adjusted for Fall 2002 Demographic Characteristics

While an intact randomized sample and complete outcome data ensure that no systematic biases enter into the simple difference-in-mean estimates of Head Start's impact, more sophisticated analysis methods provide further advantages. In addition to assignment to the Head Start study group, other factors such as a child's background and family characteristics may influence her/his outcomes in later months. If these factors can be included in models that "explain" child outcomes as the joint result of Head Start access and demographic background characteristics, uncertainty about the process used to generate outcomes will decline. In addition, confidence in the role of measured factors, including assignment to the Head Start group, will increase. This effect, known statistically as "reducing variance," will increase the chances of detecting as statistically significant any impact Head Start has on the outcomes of interest. Correspondingly, this study will be able to detect smaller impacts with 80 percent certainty, known as "minimum detectable effects," as additional factors are taken into account. This makes the research more capable of detecting Head Start impacts should such impacts occur.

---

[8] The coefficient on the dummy variable in this specification provides the impact estimate and is computationally identical to the simple difference-in-means estimate. Its statistical significance is tested using the same estimate of variance (i.e., standard deviation) as the equivalent difference-in-mean estimate and the same Student's t distribution.

[9] Impacts are initially estimated in terms of log-odds ratios, which become probabilities when passed through the logistic transformation. If assignment to the Head Start group has a statistically significant impact on the log-odds ratio when tested using the usual maximum-likelihood test procedure of a logistic model, one can conclude that it also significantly influences the probability of the outcome in question.

To add the explanatory power of background factors to the analysis, the regression models used to obtain difference-in-means estimates can be extended to express outcomes (or, in the case of logistic models, the probability of a particular outcome) as a function of both assignment to the Head Start group (the dummy variable used previously) and a set of key demographic variables measured in fall 2002. This regression equation includes a constant, the dummy variable modeling assignment to the Head Start group, and a set of key background variables. Appendix 4.3 shows this extension of the formal mathematical model underlying all the impact analyses.

The background variables used were selected in five stages, starting with a focus on the four different outcome domains (cognitive, social-emotional, health, and parenting) but then coming together into a single set of variables:

- Specification of the likely predictors of child and family outcomes for each domain, based on past research and the set of child and family measures collected by the study in fall 2002.
- Merger of the four sets of predictors (one for each outcome domain) into a single comprehensive list.
- Identification of any covariate whose role in the regression equations is at times unstable and whose coefficient, therefore, cannot always be estimated.[10]
- Removal of the unstable covariates from all regressions.[11]
- Removal (from all regressions) of covariates whose values in either age group may have been affected by the group to which a given child was randomly assigned, Head Start or non-Head Start (see next subsection).

These steps resulted in a single uniform set of covariates included in all the impact regressions that take account of child and family demographic characteristics, a list provided in Exhibit 4.2. Each demographic variable used is posited to relate to the outcomes[12] in linear fashion. For background variables that provide two-way categorizations of all the children in the sample (the great majority), this reduces to a simple shift in the average outcome level between the two groups.

---

[10] Unstable coefficients arise for a variety of reasons, most often because a two-way categorical variable has very few—or, for the replicate subsamples used to calculate variances for all the regression coefficients, no—observations in one of its cells. The SUDAAN estimation procedure used for all impact regressions could not produce numeric values of the desired coefficients in these instances.

[11] Removal of unstable covariates resulted in the consolidation of certain 5-, 7-, and 8-way categorizations of children/families as sets of dummy variables into 2- and 3-way categorizations, represented in the regressions by 1 or 2 dummy variables in each instance (omitting one of the consolidated categories each time). This collapsing of categories was necessary for child race/ethnicity, mother's education, primary caregiver's self-reported health status, and family income range in order to get all impact regressions involving demographic covariates to converge to estimated equations that contain no missing coefficients.

[12] Or, in the case of the logistic models, to the log-odds ratio.

Exhibit 4.2: Fall 2002 Demographic Variables Included in the Statistical Models Estimating the Impact of Head Start

Child Covariates

- Child Gender
- Child Age in Months as of 9/1/02
- Child Race/Ethnicity, Black (all models except for cognitive outcomes for the Spanish-English language group, and logistic models of parenting and health outcomes)
- Child Race/Ethnicity, Hispanic
- Child Has Special Needs

Parent Covariates

- Caregiver Depression Scale
- Primary Caregiver's Age as of 9/1/02
- Both Biological Parents Live with Child
- Biological Mother Is a Recent Immigrant
- Mother's Highest Level of Educational Attainment
- Primary Caregiver's Self-Reported Health Status
- Parents Are Separated or Divorced
- Mother Had a Birth as a Teenager
- Caregiver's Locus of Control Scale

Household Covariates

- Grandparent Lives in the Household (all models except for cognitive outcomes for the Spanish-English language group, and logistic models of parenting and health outcomes)
- Number of Household Moves in Last 12 Months
- Household Monthly Income Range
- Household Receives TANF

Outcomes Adjusted for Initial Fall "Starting Points"

Another set of factors helps explain child outcomes and increase the precision of the estimated impacts of Head Start: the initial fall starting points for the key outcome measures used in the impact analyses. A child's cognitive abilities measured at the beginning of her or his Head Start enrollment strongly predicts her or his cognitive abilities at the end of a year in the program (or in the non-Head Start comparison group). For this reason, a third set of the Head Start impact estimates was calculated, adjusting for each child's initial fall 2002 value on the respective outcome measures used to calculate the spring 2003 impact estimates. Thus, for example, to better explain Head Start's impact on the spring 2003 PPVT-III (adapted) scores, the fall 2002 measure of the same cognitive assessment measure (in this case each child's fall 2002 PPVT-III (adapted) score) was added to the regression analyses discussed above. Appendix 4.4 provides the particular fall 2002 cognitive assessment score, or social-emotional, health, or parenting indicator, added in this fashion for each of the spring 2003 outcomes for which impacts are examined.[13]

[13] As noted previously, the estimation procedures for including covariates in the formal regression model are described in Appendix 4.3.

There is no question that spring outcomes are, on average, higher for children who tested higher on that measure in the fall and lower for children who tested lower in the fall. Similarly, children who engaged in a particular type of behavior in the fall, or parents who adopted certain child-rearing practices in the fall, were more likely to do so the following spring. This makes the pre-test version of the outcome variable especially helpful in explaining outcomes observed in the post-test period of spring 2003, thus obtaining more precise measures of Head Start's impact in the later period. Controlling for pre-test levels of spring outcomes may also remove potential differences between the Head Start and non-Head Start samples due to nonresponse in the spring data collection. While nonresponse adjustment to the analysis weights was used to offset differential response rates, including pre-test measures as covariates helps to offset any remaining difference.

But adjustment for each outcome measure's starting point using pre-test data creates some ambiguity in interpreting the resulting impact estimates. Any differences in initial fall pre-test measures between the Head Start and non-Head Start groups will be statistically controlled when the fall 2002 outcome measures are added to the spring 2003 impact analyses that previously contained only demographic characteristics of children and families.[14] The ambiguity arises in deciding whether controlling for the initial differences in fall pre-test measures in this way enhances or diminishes the reliability of the impact estimates.

A good deal is at stake in this assessment since (based on a procedure described below) as many as 10 of the 27 pre-test measures considered as possible adjustment factors may have differed to an important extent between the Head Start and non-Head Start groups in the 3-year-old cohort, and as many as 12 of 27 measures in the 4-year-old cohort.

[14] The measure of program impact from the regression models—the coefficient on the variable indicating membership in the Head Start group—will include only that portion of the overall difference between the two groups in spring 2003 that is not accounted for by other variables in the model. Fall measures that are systematically higher (or, for factors that Head Start participation might reduce such as parental use of physical discipline, lower) for the Head Start group than the non-Head Start group and that predict child-by-child variations in spring outcomes to some degree will account for some of the systematically higher (or lower) spring outcomes for the Head Start group, precluding the coefficient measuring program impact from doing so.

Whether one should adjust for such factors depends on **the reason** the fall 2002 measure of the spring 2003 outcome differs systematically between the Head Start group and non-Head Start group. In many randomized impact studies, removing initial pre-test differences on outcome measures between the intervention group and the control group can only enhance the impact estimates. Specifically, initial differences on average fall outcomes measures between the two groups may occur as a result of one or more of the following reasons:

- chance differences between the types of children put in the two groups during random assignment;
- corruption of the random assignment process that undercuts the intention of giving every child an equal probability of being randomly assigned to the Head Start group regardless of his or her characteristics or any other individual-specific factors; or
- omission of different types of children from the Head Start group's spring 2003 analytic sample from those in the non-Head Start group's spring 2003 analytic sample due to potential differential nonresponse rates between the groups in collecting the spring outcome data.[15]

If one or more of these factors contributes to the observed initial differences between groups on the initial fall outcome measures, then adjustment for this initial difference in fall outcome scores will improve that impact estimate. If, on the other hand, the initial differences on fall outcome scores result from an early impact of the Head Start intervention, then the inclusion of the initial fall scores in the model may attenuate the resulting impact estimates. In this instance, Head Start does not get "full credit" for the impacts it achieves by the time outcomes are measured in spring 2003 since some of those potential early impacts—the portion that may have occurred by the time the initial fall 2002 data were collected—are not counted. These potential early impacts will be removed from the spring impact estimates by the inclusion of the initial fall outcome measures in the analyses.

In many randomized impact studies, the pre-test versions of the important program outcome variables, as well as all demographic characteristics, are measured **prior to, or at the point of, random assignment,** when one can be sure that differences between the intervention and control groups (if any) do not reflect early impacts of the program.[16] In those instances,

[15] Differences in average fall "outcomes" between the two groups would arise in this instance only to the extent that the analysis weight adjustments for dealing with nonresponse described earlier do not completely compensate for differential nonresponse.

[16] Regardless of the intervention, or the channels through which it might influence the initial true characteristics of sample members or the way those characteristics get reported to the evaluation, that influence cannot possibly begin for one group (but not the other) when no one knows which children are in which group. This is precisely the situation that must hold prior to random assignment for the Head Start applicant families and the grantees operating the program (and even for members of the evaluation team collecting the data).

removal of any measured differences between the two groups in pre-test values or demographic characteristics when calculating subsequent impacts necessarily improves the estimates, whether those differences are caused by chance, corruption of random assignment, or differential nonresponse during data collection. So where feasible, experimental evaluations measure all sample members' characteristics prior to randomization and adjust for them in the impact analysis without fear of potentially doing harm to the estimates. Two of the demographic variables used in this analysis (see Exhibit 4.2) fit into this category: child gender and race/ethnicity. Both come from rosters completed by program intake staff prior to random assignment.

Unfortunately, data collection prior to random assignment was infeasible for many of the important demographic variables in Exhibit 4.2 and all the initial developmental and behavioral "starting point" measures listed in Appendix 4.4. When (1) in-depth in-person information must be collected for a large, highly geographically dispersed experimental sample—implying high costs of data collection per sample member—and (2) notification of acceptance into the program must occur quickly once eligible applicants are identified, most background data cannot be collected prior to random assignment. Additional data collection for persons ultimately found ineligible for inclusion in the program or the research would be unethical and inordinately costly. These were exactly the circumstances of the National Head Start Impact Study when random assignment took place in mid-2002.[17]

Due to these constraints, most of the fall 2002 data on children and families in the study were collected over a 3-month period from October 2002 through December 2002 (with most completed by mid-November) at a considerable lag from random assignment. As a result, the possibility that some early Head Start impacts may have preceded fall 2002 data collection for many children cannot be ruled out. Moreover, these potential early impacts could account for some, or all, measured differences in characteristics between the Head Start and non-Head Start samples at that point. There are two exceptions, however, among the demographic variables

[17] In-depth in-person data collection was made necessary because of the number and complexity of the child assessment scales needed for each child, which could only be administered in person by highly trained staff. The high unit cost of this type of data collection dictated that the minimum number of children be put in the sample and assessed, putting a premium on identifying children certain to be included in the evaluation before initiating field data collection. Thus, data collection could not begin until a firm determination was made that a particular child would be randomly assigned into the study sample. This required that the Head Start provider organization deem the child appropriate for services based on its local service-targeting priorities. However, at the point this determination was made, the grantee also faced substantial pressure to notify families selected for the Head Start group that their children would be allowed to participate in the program. This forced random assignment to take place—and in many cases actual Head Start program participation to begin—almost immediately after eligibility was determined, leaving no time for baseline data collection prior to that point. Postponing random assignment long enough for extensive in-person testing of children to take place would have imposed an unacceptable hardship on families and Head Start agencies left wondering which children would be served by the program. Accelerating data collection to substantially precede eligibility determination would inevitably have led to many costly interviews and assessments being conducted for children and families who in the end proved ineligible for inclusion in the study.

measured after random assignment: whether the biological mother of the child in question was a teen parent and whether she first arrived in the US within the last 5 years. We do not believe these measures could have been affected substantially by the program[18] and, hence, have included them in all analyses involving demographic background factors. All the remaining demographic variables in Exhibit 4.2 (other than those noted earlier as collected prior to random assignment), including measures of living arrangements, marital status, income, educational attainment, and the health status of the child's primary caregiver, in theory could have been influenced by the Head Start intervention prior to measurement in fall 2002. This same concern arises for the fall 2002 measures of developmental and behavioral "starting points" listed in Appendix 4.4. Any potential early impact of Head Start on these variables will be excluded from spring impact estimates when the analysis controls for pre-test measures when calculating effects.

To make the decision of which control variables ("covariates") to include in the analyses, a statistical procedure was developed for this study that tests whether appreciable early impacts on factors measured in fall 2002 could have occurred. Rather than presume that no such impacts occurred unless the data prove otherwise (as one would do if the usual test for statistical significance were used), the procedure adopted requires strong evidence that early impacts of an appreciable magnitude **did not occur.** Only then does the tradeoff between possible small omissions of potential early fall impacts from the spring impact estimates on the one hand, and gains in statistical precision plus removal of nonresponse bias on the other, become favorable and warrant the inclusion of the initial fall outcome scores in the analyses.

The procedure adopted seeks a 90 percent assurance that Head Start's potential early impact on fall demographic characteristics, and on initial fall outcome measures, was small or nonexistent.[19] In such cases, the risk of potentially excluding a small portion of the overall impact from the spring 2003 impact estimates is more than offset by expected precision gains and nonresponse bias reductions from including the initial fall outcome variables in the analyses.[20]

---

[18] A biological mother of a child applying for Head Start could not have become a parent for the first time following random assignment, so teen-parent status for all mothers was already established well ahead of the point where Head Start participation began and could have had an effect on actual fertility. Similarly, with all families in the research sample living in the U.S. at the time of application, random assignment could not have changed the fact or timing of immigration.

[19] Appendix 4.5 describes the procedure used. "Small" is defined on a relative basis (an effect size of 0.2 or smaller) that takes account of how much the fall measure varies in the population being studied using a guideline suggested by Cohen (1988) that keys off effect size (the ratio of impact to standard deviation, a measure of variation). See Jacob Cohen. (1988). ***Statistical Power Analysis for the Behavioral Sciences***. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

[20] This is true notwithstanding prior steps to remove a portion of the nonresponse bias, if any, through sample weight adjustments. These adjustments, described in Appendix 4.1, remove only those differences attributable to a subset of the fall background characteristics one might want to take into account. By stipulating that the influence of all background factors is approximately linear, impact regressions can adjust for many more (actually, an almost unlimited number of) fall measures that could differ between the

Otherwise, it may be unwise to adjust for the initial fall outcome scores in computing spring impact estimates.

Demographic characteristics that failed the test, other than those noted above as appropriate for inclusion on other grounds, were simply omitted from the impact analyses presented in each of the following chapters. Impact estimates computed both with and without the inclusion of the initial fall outcome measures are presented in the tables within each chapter.[21] However, the actual discussion of the relevant impact findings highlights the impact estimates that, in our view, provide the best evidence of Head Start impact. Based on the test for appreciable initial differences in fall measures, when strong evidence is found that Head Start exposure has at most a small effect on the fall pre-test measure, the discussion favors the impact estimate that includes the fall outcome measure as a covariate, with an anticipated gain in precision and nonresponse bias removal as a result. When such evidence is not available, a cautious approach is taken emphasizing impact estimates that exclude the initial fall outcome measure as an explanatory variable.[22] This avoids all risk of excluding part of Head Start's overall impact when reporting spring findings.

Presentation of Results

Tables of results presented in subsequent chapters include all three perspectives for measuring Head Start's impacts discussed above, i.e., (1) simple differences in average outcomes, (2) impacts adjusted purely for demographic characteristics that are clearly unaffected by random assignment to the Head Start or non-Head Start group, and (3) impacts adjusted for both demographic characteristics and fall 2002 developmental/behavioral starting points. Based on the potential risks and rewards of adjusting impact estimates for differences in fall 2002 pre-test outcome measures (as discussed above) for each statistically significant impact, the tables highlight the single most appropriate measure among the three. Overall conclusions of the

---

Head Start children in the analysis sample and the non-Head Start children in the analysis sample due to differential nonresponse in the spring. The regressions are not constrained by the rapidly shrinking cell sizes that typically limit the number of factors that can be taken into account through stratified matching and reweighting of the data.

[21] Language of assessment in the fall (Spanish, English, Other) affected the way impact estimates were calculated for the combined analysis sample. Three different versions of the impact estimates were computed for the following outcome measures: PPVT-III adapted, CTOPPP Elision, WJ-III Oral Comprehension, WJ-III Spelling, Letter Naming Task, and WJ-III Applied Problems. For each outcome, different impact estimates were derived using as fall 2002 covariates: (i) "English PPVT-III adapted" and "Spanish PPVT-III adapted", (ii) "English PPVT-III adapted" only, or (iii) neither language-specific PPVT-III variable. A similar specification was used to estimate impacts on WJ-III Letter-Word Identification scores using the language-specific versions of this test in fall 2002. The exhibits used to present findings in Chapters 5 through 8 present the results of versions "ii" and "iii" with version "i" discussed in a footnote to the tables.

[22] For instances in which three different versions of the impact equations were estimated (see Footnote # 21), results favor the version of estimated impact that includes "English PPVT-III adapted" from fall 2002 (or its WJ-III Letter-Word Identification score equivalent) as a covariate. There is strong evidence that Head Start exposure has at most a small effect on this measure for the sample assessed primarily in English while such evidence is lacking for children assessed primarily in Spanish.

research, including the summary of findings presented in the Executive Summary, draw from findings on the preferred measure only; in practice, the pattern of results does not differ much across the three approaches. For context, the tabular results also include the average outcome levels for the Head Start and non-Head Start samples.

The discussion of the preferred findings also provides the corresponding effect sizes, which are defined as the impact estimates divided by the standard deviation of the outcome measure in the population, providing a "yardstick" for gauging the quantitative importance of a measured impact in relation to the natural variation of the child or family outcome Head Start is seeking to affect. Many researchers have used Cohen's (1987) guidelines for interpreting the relevance of effect sizes, with an effect size of 0.2-0.5 being considered small, 0.5-0.8 moderate, and over 0.8 is large.[23] Within the field of education research, some researchers have argued that an effect size has to be at least 0.25 or 0.33 of a standard deviation to be considered "educationally meaningful" (Slavin, 1990; Wolf, 1986).[24, 25]

In contrast, Glass et al. (1981)[26] and McCartney and Rosenthal (2000)[27] have asserted that the effect sizes derived from a given study always should be interpreted within the context of the empirical literature on comparable interventions designed to produce similar effects. In the NICHD Study of Early Child Care, the quality of child care predicted children's cognitive performance at 54 months (range of effect sizes was 0.04 to 0.08).[28] The Tennessee study examining the benefits of smaller class sizes in the early school grades yielded effect sizes that ranged between 0.13 and 0.27 on several direct assessments of children's reading and math performance (Finn & Achilles, 1990).[29] A meta-analysis of evaluations of family support programs yielded the following weighted mean effect sizes across several key outcome domains: children's cognitive development (0.253), social-emotional development (0.258), physical health and development (0.091), parenting attitudes and knowledge (0.182), parenting behavior (0.246),

---

[23] Cohen, J. (1987). ***Statistical Power Analysis for the Behavioral Sciences*** (2nd ed.). Hillsdale, NJ: Erlbaum.

[24] Slavin, R.E. (1990). ***Cooperative Learning: Theory, Research, and Practice***. Englewood Cliffs, NJ: Prentice Hall.

[25] Wolf, F.M. (1986). ***Meta-analysis: Quantitative Methods for Research Synthesis***. Newbury Park, CA: Sage.

[26] Glass, G.V., B. McGaw, and M.L. Smith. (1981). ***Meta-Analysis in Social Research***. London: Sage.

[27] McCartney, K. and R. Rosenthal. (2000). "Effect size, practical importance, and social policy for children." ***Child Development***, Vol. 71(1), pp. 173-180.

[28] NICHD Early Child Care Research Network & Duncan, G. (2003). "Modeling the impacts of child care quality on children's preschool cognitive development." ***Child Development***, 75(5), pp. 1454-75.

[29] Finn, J.D. and C.M. Achilles. (1990). "Answers and questions about class size: A statewide experiment." ***American Educational Research Journal***, 27(3), pp. 557-577.

and family functioning/family resources (0.284) (ACF, 2001).[30] Finally, another recent meta-analysis of 33 studies focusing primarily on early childhood education programs for low-income 3- and 4-year-olds revealed a weighted mean effect size of 0.118 across the studies reviewed (Aos, Lieb, Mayfield, Miller, & Pennucci, 2004).[31] Reflective of their contextual basis for judging impacts to be important in magnitude, this report uses the following convention: less than 0.2 is small; 0.2-0.5 is moderate; and greater than 0.5 is large. This allows interpretation of effect sizes within the broader context of findings from other similar early childhood intervention studies.

Analysis of Subgroups and Moderating Factors

To this point, the discussion has focused on ways to measure the impact of Head Start on the *average* child or family in the program. Of course, impacts will likely vary across different subsets of the children and families served. For example, Head Start may benefit boys more than girls (or the reverse), or it may benefit families headed by a single parent more than two-parent families (or the reverse).

In addition to an interest in the overall national impact of Head Start on children's school readiness, Congress mandated an examination of how impacts vary for different types of children and families. The intent is to understand "what drives the overall impacts" when the program is having an effect of important magnitude for the average Head Start participant. In particular, there is interest in determining the extent to which the benefits of Head Start may be widespread—i.e., whether the benefits reach many types of children and families to produce the overall average effect rather than benefiting some but having little or no effect on others.

Identifying groups of children (or families) that benefit more or less from Head Start may have important policy and program implications. It can suggest areas where the program needs to be strengthened or enhanced to ensure that all participants advance in their development. For example, Head Start programs are required to serve children with special needs so it is important to understand the extent to which these children benefit from their participation over and above an interest in determining if Head Start improves the lives of the average participant. In addition, prior early childhood research has indicated that some groups of children follow different

[30] Administration for Children and Families. (2001). ***National Evaluation of Family Support Programs: Final Report***. Washington, DC: Author.

[31] Aos, S., R. Lieb, J. Mayfield, M. Miller, and A. Pennucci. (2004). ***Benefits and Costs of Prevention and Early Intervention Programs for Youth***. Document #: 04-07-3901: Washington State Institute for Public Policy.

developmental paths and may, as a consequence, be assisted by Head Start in distinctive ways, such as children in racial and ethnic minority groups and non-English speaking children and parents.

This interest in "who benefits?" motivates two types of analyses. The first considers the impact of Head Start on **individual subgroups** of program participants, asking for example: Does Head Start help Hispanic children? Children with single parents? Immigrant families? Mothers who first gave birth as teens? Special needs children? This same set of results can be considered in total to determine whether certain subgroups "drive" the overall average impact or whether widespread benefits accrue to many different subgroups.

The second set of analyses considers whether impacts differ in magnitude **between** distinct types of children and families. For example, Head Start may have smaller effects on children of recent immigrants than on other children or larger effects on two-parent families than single-parent families. Interest in these comparisons stems from several sources:

- Researchers want to know what factors "moderate" the influence of early childhood services (such as those provided by Head Start) on child development and family functioning. In this case, the term "moderate" means alter the size of the impact of those services when they are provided to one type of child (or family) versus another. For example, the extent to which a child's primary caregiver reports symptoms of depression may moderate how much Head Start is able to help him/her develop good social skills, or a child's home language may moderate the program's ability to expand reading readiness by getting parents to read more to their child.

- As noted above, Congress required that the study identify the types of children and families that benefit **most** from Head start participation, a question that implicitly relates impacts for one type of child/family to impacts for another. For example, do younger children benefit more than older children? Single parent families more than two parent families?

- Head Start program operators might seek to enhance services in ways that would particularly benefit subgroups found to be experiencing smaller impacts than other subgroups, such as children with special needs or families with diverse cultural or linguistic backgrounds.

With sufficient data, all subgroup impacts and moderator influences would become apparent when the difference in outcomes between Head Start and non-Head Start families in one subgroup is calculated and compared to the difference in outcomes between Head Start and non-Head Start families in another subgroup. But because data are limited, the study cannot decisively answer all questions about Head Start's impact on different subpopulations. Still, where evidence is strong that an impact on a particular subgroup, or a difference in impacts between subgroups, has occurred, the subgroup analysis will produce a finding that is not difficult to interpret: real impacts in the measured direction have taken place.[32] In contrast, a non-significant finding is more ambiguous and could indicate either: (1) that there is in fact no impact, or difference in impact, for some subgroup(s) or (2) that impacts exist but are too small in magnitude to reach the threshold of what the data are able to detect. This means that statements about the subgroup and/or differential impacts that did take place will be much less equivocal than statements about impacts being lacking or undifferentiated between groups. The latter ambiguity makes it hard to be conclusive about which subsets of children and families "drive" the overall results (since additional subpopulations may contribute but escape detection), judge impacts to be widespread as opposed to narrowly concentrated, or identify subgroups that have similarly sized impacts.

As a consequence, this preliminary examination of subgroup impacts will, at times, seem incomplete. But the reader should keep in mind that the goal here is to present all the evidence available in the data on subgroup-related questions. Each piece is valid in its own right; yet the full picture may be a patchwork due to statistically inconclusive, ambiguous findings for many subgroups and potential moderating influences.

Exhibit 4.3 lists the subgroup-defining (i.e., moderating) factors examined in the current analyses across the four outcome domains of interest: cognitive, social-emotional, health, and parenting. All subgroups and moderators considered here were identified in advance of the data analysis on the basis of their program and policy importance to Head Start or their relevance to understanding early childhood development generally. Different subpopulations are germane to

---

[32] Strong evidence of a subgroup impact or difference in impacts takes a more complex form here than usual. Because so many different subgroups and subgroup differences are tested, at least some will appear to be statistically significant by chance alone. This follows from the construction of standard tests of statistical significance for individual results. Because not all uncertainty can be removed from the analysis of sample-based data, individual tests of statistical significance must be constructed to allow the possibility of an incorrect conclusion on some occasions—typically, a 5 percent chance. Thus 1 in every 20 cases in which no impact (or difference in impact) has occurred will produce a statistically significant finding. When, in the face of no actual impacts, many tests are run for impacts by subgroup or between subgroups, some of the tests are certain to have this feature—i.e., they will produce "false positive" results. One has no way of knowing which, if any, of the potentially many statistically significant results on subgroups or subgroup impact differentials constitutes a 1-in-20 "false positive." Hence, testing for whether a "false positive" exists among many statistically significant results (a statistical finding that would itself carry some uncertainty) is of no value absent the ability to determine which one it is. More useful would be a test of whether **all** significant findings on subgroups are "false positives"; until this possibility is ruled out, one should not draw strong conclusions from subgroup analysis. Two procedures are used here for ruling out false positives. Subgroup impacts for a particular outcome measure such as oral comprehension or use of dental care cannot all be "false positives" when overall impact on that outcome is significant, since **some** subgroup must have benefited if the average child or family did. In addition, the set of significant impacts for a particular subgroup (or for the difference in impact between two subgroups) is very unlikely to consist of only "false positives" when an important share of all the impacts tested for that subgroup (or of all difference in impacts tested between two subgroups) are individually significant) Adopting a cautious approach for this purpose, the analysis in later chapters requires that three times the share of significant results predicted by chance alone when no true impacts occur be significant in order to consider the whole set of results real, i.e., that at least 15 percent of all tests run for a given subgroup or subgroup comparison be statistically significant at the 95 percent confidence level. When this standard is met, or the preceding standard based on a significant average impact for the full sample, it is appropriate to consider each individual subgroup finding reliable in its own right.

different domains of outcomes on this basis, cognitive, social-emotional, health, and parenting. Within a given domain, all subgroups are tested for a common group of outcomes (the set used for the overall impact analysis) so that findings, both statistically significant and insignificant, can be assessed as a group to determine how widespread the benefits of Head Start are and identify within this pattern child/family types that definitely benefit.[33] The reasons for selecting the particular subgroups and moderators in the exhibit are discussed below:

- **Whether the child has special needs.** Parents reported whether their child had one or more special needs (e.g., learning disability) as of fall 2002. Evidence from Head Start FACES and other studies indicates that children with special needs have lower cognitive scores and lower levels of social skills and higher levels of problem behavior at the beginning and end of the program year, especially after controlling for differences in family circumstances. This could lead to either statistically significant differences in effect sizes or significant impacts on one of these two subsets of children, special needs and non-special needs, and not the other, or both of these patterns. In addition, the existence and size of the health impacts that Head Start can achieve may also be influenced by the presence of disabilities. Consequently, it is important to learn whether the benefits that children with disabilities derive from the opportunity to participate in Head Start are similar to or greater than those of children without disabilities, and whether benefits exist at all in each instance.
- **Child race/ethnicity.** Children were categorized as African-American, Hispanic, or White/Other (which includes Asian and Native American). Close to two-thirds of all children enrolled in Head Start are from racial and ethnic minority groups, especially African-Americans and Hispanics. These children, on average, enter Head Start with relatively lower cognitive scores, lower levels of parent-reported social skills, and higher levels of parent-reported problem behavior, even when lower parent education and family income levels are taken into consideration. In addition, many minority children enter Head Start with greater health needs than non-minority children and may not have the same opportunities to obtain health, vision, and dental services. Finally, there is some evidence of the use of harsher discipline by low-income African American parents.[34] As a result, one might expect the impact of Head Start to be greater for minority children than for White children from low-income families. Their differences, as well as differences in impact between Hispanic and African-American children, will be important to detect, as will the existence of impacts for each one of these racial/ethnic groups in its own right.
- **Child gender.** Previous research has found significant gender differences in the frequency of cooperative, pro-social, and problem behavior among young children. For example, aggressive and hyperactive behaviors tend to be more common

[33] The word "definitely" here as its usual statistical meaning of something that is true with a very high degree of certainty—95 certainty in this case. Some such conclusions will be wrong, 1 in 20 on average, but these constitute "false positives" unavoidable in any statistical analysis. As long as one does not add subgroups to be examined during the course of the analysis, a process guaranteed to produce a statistically significant "false positive" finding eventually, mistaken conclusions on who benefits, each taken on its own terms, continue to be very unlikely events.

[34] Pinderhughes, E.E., K. Dodge, J. Bates, G. Pettit, and A. Zelli (2000). "Discipline responses influences of parents' socioeconomic status, ethnicity, beliefs about parenting, stress, and cognitive-emotional process." ***Journal of Family Psychology,*** 14(3), pp. 380-400.

among boys than girls, whereas there is less of a gender difference with respect to withdrawn or depressed problem behavior. Some evidence indicates that raising young boys may present more challenges to parents than raising young girls.[35] Given these rate differences, it is of both theoretical and practical relevance to ask separately whether Head Start conveys benefits for males and for females, and whether the size of impacts differ between the two groups.

- **Language of child assessment.** Children were assessed in either English in both fall 2002 and spring 2003, or in Spanish in the fall and English in the spring. Hispanic children from Spanish-speaking families are one of the fastest growing segments of the Head Start child population. But many local programs have to struggle to provide staff that are fluent enough in Spanish to communicate easily with both children and parents, as well as to assist children in their acquisition of English. It is important, therefore, to learn whether the benefits that these children derive from Head Start are equivalent to, less than, or greater than those of children from other language backgrounds, and whether impacts occur at all for each of the language groups in its own right.

- **Child's home language.** Parents reported the language that was most often spoken at home, which was categorized as English or not English (this variable was used as a moderator for health and parenting outcomes to capture the language of the parent rather than the child). Families whose home language is not English may be harder to engage in efforts to improve their parenting skills and may not be tied into the social welfare safety net as much as those whose home language is English, due to language and cultural barriers. As a consequence, it is important to learn whether children from non-English-speaking homes benefit from the opportunity to participate in Head Start and if those gains differ from Head Start induced impacts in English-speaking homes.

- **Parent's marital status.** Single-parent families and families with parents who are separated or divorced may be unable to provide their children with as much stability and parental resources as married-parent families, leading to greater needs in the children (as well as emotional and behavioral issues). For example, previous research has indicated that children from separated, divorced, and unmarried family situations have higher levels of difficulties in elementary school, compared with children from stable married families. In addition, Head Start emphasizes parental involvement, and parents from married-parent families may be more able to take an active part in the Head Start program, resulting in greater benefits to their children. It is, therefore, important to learn whether the impact of Head Start is different for children from different types of home environments, and to test separately for impacts in households of each distinctive marital arrangement.

- **Primary caregiver's depression rating.** As part of the fall 2002 interview, parents also reported on the CES-D scale for depressive symptoms.[36] A frequent occurrence within low-income families, especially in single-parent families, is depression in the child's primary caregiver, typically, the mother. Such instances of depression may pose an obstacle to the parent's participating in Head Start as much

---

[35] Leaper, C. (2002). "Parenting Girls and Boys." In M.H. Bornstein (Ed.), ***Handbook of Parenting, Vol. 1, 2nd Edition.*** Hillside, NJ: Erlbaum.

[36] Ross, C.E., J. Mirowsky, and J. Huber. (1983). "Dividing work, sharing work, and in-between: marriage patterns and depression." A***merican Sociological Review, 48***, 809-823. For this analysis the continuous scale scores were used rather than clinical cutoff scores for depression.

as is optimal and in providing support for the child's learning, social-emotional development, and health care needs. Parents who are struggling with mental health issues may also be less receptive to efforts to strengthen their discipline practices and to increase their engaging in various educational activities at home with their child. Consequently, the impact of Head Start may be less in situations of high levels of caregiver depression at baseline. On the other hand, to the extent that Head Start is a compensatory program designed to make up some of the support and resources that the home is not providing, children with depressed primary caregivers might be expected to show greater benefits from Head Start. The program may also assist the depressed parent in obtaining the necessary services and supports for his/her depression and thus indirectly benefit the child's social development and emotional well-being. For all these reasons, an important question is whether degree of depression moderates the size of Head Start's impact.

- **Mother's age at first birth.** Parents were asked, "How old were you when you gave birth for the first time?" as part of the fall 2002 parent interview. Mothers were divided into two groups, those who had given birth before or after age 19 (referred to as "teen" versus "non-teen" mothers), and this variable was used as a moderator for parenting outcomes. Mothers who first gave birth as adolescents are at greater risk for poor childrearing practices. However, because of their heightened risk, Head Start may make special efforts to engage them in services that may provide important benefits more than for mothers who were older when they first became parents.

- **Child's achievement at the start of Head Start.** The child's score on the outcome variable as of fall 2002 was included as a moderator for both cognitive and social-emotional outcomes. FACES has repeatedly found that children who enter Head Start with lower cognitive scores (e.g., those in the lowest quartile of the Head Start child distribution) show larger cognitive gains from fall to spring than the children with average or above-average entering scores. This phenomenon has been interpreted as indicating that Head Start is of particular benefit to children with larger cognitive deficits. Others have argued that the finding is merely a manifestation of "regression to the mean," wherein those who are lagging at a particular point of measurement make the largest advances simply by moving back closer to the middle of the distribution. Since this is the natural tendency in a developmental process in which children show spurts of growth at different times, it is not necessarily an indication that Head Start is especially effective for those with greater initial deficits. The crucial issue is whether the progress of children with lower initial achievement is further sped by participation in Head Start, precisely what comparison of the progress made by the Head Start and non-Head Start group children at lowest levels of initial achievement will illuminate. If true program-created impacts have occurred, these cognitive gains may also have positive carryover to social and emotional development, making it important to also look at outcomes in those domains for children with the lowest initial cognitive scores.[37]

[37] In this comparison, any "regression to the mean" that occurs for the Head Start sample will be matched and hence cancelled out by similar regression in the non-Head Start sample.

Exhibit 4.3: List of Variables Used As Moderators by Outcome Domain

Outcome Domain	Moderators
Cognitive	Child Has Special Needs Child's Race/Ethnicity Caregiver Depression Language of Child Assessment Caregiver Married Fall Measure ((PPVT, Bear Rate, Draw Score (3-year-old group only), and Color Score (3-year-old group only))
Social Emotional	Child Has Special Needs Child's Race/Ethnicity Caregiver Depression Language of Child Assessment Caregiver Married Caregiver Separated or Divorced Child's Gender PPVT
Health	Child Has Special Needs Child's Race/Ethnicity Caregiver Depression Caregiver Married Home Language
Parenting	Child's Race/Ethnicity Caregiver Depression Language of Child Assessment Caregiver Married Child's Gender Was Mother a Teen at First Birth? Home Language

This set of moderator variables defines subgroups for which separate impact estimates can be calculated. It is important, therefore, to ensure that each moderator and each subgroup determined by a categorical moderator is independent of the intervention. This is so that subgroups are well-matched between children originally randomized into the Head Start or non-Head Start groups and represent valid experimental analyses in and of themselves. For example, if Head Start participation led to greater awareness on the part of parents of a child's special needs before the fall 2002 data collection, comparisons of the children parents report as special needs in the Head Start group versus the non-Head Start group would not be based on a consistent, matching set of individuals. This could bias measures of impact in this population.[38] To avoid this risk, the same evidence of **lack of early program impact** was required on each moderator variable used (and in both age groups) as previously required of covariates for the

[38] Continuing the example, the children identified by parents in the Head Start group might have less severe needs than other special needs children and hence better developmental outcomes in spring 2003 data. Their inclusion in the Head Start portion of the special needs subgroup analysis, but not in the comparison group portion, would lead to a misleadingly favorable estimate of the program's impact. Their absence from the non-special needs subgroup analysis would produce an inappropriately unfavorable indication of Head Start's impact in that subpopulation.

regression analysis. Indeed, many of the moderators and subgroup definers in Exhibit 4.3 are the same as the covariates described earlier (compare to Exhibit 4.2 above). All the others listed here were either measured prior to random assignment (e.g., home language) or have also been convincingly shown to have effect sizes that are at most judged to be small by the statistical standard used (e.g., mother's marital status).

A single regression provides information on both topics of interest: how impacts vary with the moderating factor examined and, if that factor is a 0/1 indicator of membership in a particular group, how large an impact Head Start had on each of the subgroups defined by the moderator variable. This analysis interacts the dummy variable for assignment to the Head Start group with each moderator variable in turn, allowing impact to vary with that factor. For 0/1 moderators such as gender (e.g., 0=boy, 1=girl), Head Start's impact on individual subgroups can be inferred from the coefficient on the random assignment dummy variable (impact on the omitted subgroup, in this case boys) and the sum of that coefficient and the coefficient on the interaction term itself (impact on the included group, or girls). For moderators that indicate membership in a subgroup, this procedure replicates a difference-in-differences approach that estimates each subgroup impact as the Head Start/non-Head Start difference in mean outcomes for individuals in that subgroup and measures the effect of the moderator on the size of impact as the difference between those two estimates. Continuous moderators (e.g., parental depression) produce regressions that indicate if the impact of Head Start varies with the value of the moderating variable.

Depending on the analysis perspective adopted (see earlier discussion) various fall 2002 measures are added to the regressions as covariates. Tests of the statistical significance of both potential moderating influences and the average impact of Head Start in a given moderator-defined subgroup are derived along lines similar to those used in testing for the overall impact of Head Start. Details of the moderator and subgroup analysis regression approach and test procedures appear in Appendix 4.3.

Two special restrictions were placed on the subgroup/moderator analyses. First, to avoid findings that may exaggerate contrasts between subgroups due to the vagaries of small-sample analysis, subgroups with fewer than 50 observations in either the Head Start or non-Head Start group were not examined. Second, certain observations could not be included in particular subgroup analyses for certain moderators. For example, children with deceased parents could not be classified in examining how impacts vary with mother's current marital status and so were left out of that particular moderator regression. Similarly, children who could not be assessed in either English or Spanish in fall 2002 due to lack of familiarity with these languages were dropped from the analysis sample when examining initial cognitive ability as a moderator of impacts.[39]

Impact findings for subgroups and moderators are presented in the following chapters using two of the three perspectives introduced in the examination of overall results: mean differences adjusted for fall 2002 demographic characteristics and mean differences adjusted for fall 2002 demographic characteristics and developmental and behavioral starting points. The preferred perspective is again highlighted in the discussion for each impact or impact difference considered.

Estimating the Impact of Program Participation

All of the impact estimates described to this point measure the effect of Head Start on the average child **randomly assigned to the Head Start group.** However, as discussed in Chapter 2, not all of these children actually participated in federally funded Head Start services, the intended treatment. This is not an unexpected phenomenon: in the normal course of events, some children and families accepted into Head Start never participate, because their interest in what the program has to offer has declined since application, because other center-based arrangements have been found, or because other events interrupt plans to attend (e.g., moving to another city or distant neighborhood). This suggests two different versions of the research questions posed at the beginning of the study:

- How much does Head Start help the typical child and family **admitted to** the program, on average?
- How much does Head Start help the families and children that **actually participate in** Head Start, on average?

It will be harder to improve the average outcome of everyone accepted than the average outcome of participants, assuming that non-participants gain little or nothing from the program. If the non-participation rate (also known as the “no-show” rate) exceeds 5 or 10 percentage points, the magnitude of the difference may matter.

---

[39] Some cognitive measures were collected for all children in fall 2002 regardless of language background and could be analyzed as moderators for all sample members. These included tests administered without major reliance on a particular spoken language, such as counting bears, color naming, and the McCarthy drawing test and the PELS measure based on parent interviews.

Answers to both questions matter for policy and program administration purposes. Head Start programs are typically funded for a fixed number of slots, regardless of whether all slots are used. In that sense, the Federal program pays for slots rather than actual participants where the two differ, so impacts **per family or child admitted** has some relevance to the fiscal picture. Also, the Head Start program can offer **opportunities** to participate but it cannot compel any child to attend. Hence, the impact of admission into the program, whether taken or not, measures the typical result of what grantees do—provide access—rather than the effect of delivering services to every selected child and family.

Yet the question of how much children gain from actually participating in Head Start's services remains an important one. For local programs at full attendance (not simply full enrollment, on paper) impacts **per participant** correspond with Federal funding per slot. Moreover, if impacts per participant are large but impacts per admitted child comparatively small, the evaluation will show the value of increasing participation rates as an adjunct or alternative to expanding the number of children accepted into the program. Procedures for estimating program impacts on participants are discussed below.

An Experimentally Based Strategy of Estimation

A research study in which random assignment to the intervention group dictates **access** to program services but not actual **utilization** of those services cannot directly estimate the average impact of program participation. This is because the Head Start group includes "no-shows" who, when granted access to the program, did not actually participate. The non-Head Start group includes equivalent types of children and families who would not have participated had they been given access.

One could look at outcomes only for actual participants in the Head Start group (excluding the "no-shows"). But the subset of the non-Head Start group that corresponds statistically to these individuals cannot be identified in equivalent fashion—there is no information to identify which of the non-Head Start children would have participated in the program had they been granted access.

Fortunately, the best way to estimate Head Start's impact on the average participant **does not require that one knows anything about why no-shows arise, or how they differ from other families and children in the sample.** If it did, any impact measure produced for the participant population would have the same drawbacks that affect quasi-experimental estimates from non-randomized studies: selection bias caused by pre-existing differences between participants and comparison group members. But if one can assume that no-shows experience zero impact from Head Start, it is possible to avoid these kinds of assumptions about (or analyses of) selection into and out of the program. That is, "no-shows" can be entirely different from participants in measured and unmeasured ways, but it is unnecessary to understand how they are different or to make any adjustments for their distinctive characteristics.

This is possible by using the original comparison of **all** Head Start-group members to **all** non-Head Start group members but interpreting it in a different way. The new interpretation says that the Head Start group's impact—how its outcomes differ from what would have transpired without a Head Start program—has two components:

- The impact on "no-shows" who by definition do not participate in the program, even though admitted, which can logically be assumed to be zero.
- The impact on everyone else assigned to the Head Start group—i.e., on the Head Start participants who comprise the rest of the experimentally determined intervention group.

This assumption alone—the presumption that children and families who never receive Head Start services remain unaffected by their assignment to the program group—makes it possible to translate the measured effect of the program on the entire Head Start sample (which the experimental design provides directly using research methods described in earlier sections) as a way to assess the average effect of Head Start on just the participants.[40] It does not matter what the average effect would have been on non-participants had they participated. Nor does it matter whether non-participants have different outcomes than participants due to "selection" or pre-existing differences. Before focusing on this assumption's plausibility, the next section traces the implications for measuring in a reliable fashion the impact of Head Start on the average participant.

It is important to understand that the overall impact on all children is simply a weighted average of the impact on participants and the impact on the "no-shows":

Impact (on all children) = P (impact on participants) + Q (impact on no-shows)

[40] Appendix 4.6 discusses the basis for this assumption.

where P is the number of participants in the Head Start group and Q is the number of "no-shows" in the Head Start group. If the assumption of zero impact on the "no-shows" is incorporated into this equation, one gets the following expression:

Impact (on all children) = P (impact on participants) +
O {Impact (on all children)}/P = (impact on participants)

This does not say that the *average* effect on participants is the same as the *average* effect on the whole sample, derived from previously described analyses, when no-shows experience a 0 effect. Instead, the average effect on any set of individual children or families depends on both the **total** amount of gains accruing to **all** the individuals in the group—the measures represented by the "Impact (____)" terms above and the **number** of individuals in the group. In effect, P just rescales the total gain to all Head Start sample group members by dividing by the number of participants rather than the number of Head Start group members overall.[41]

The most important aspect of this "no-show adjusted" estimated average impact on program participants is demonstrated by Bloom (1984)[42] in the seminal article on this topic and shown to be equivalent to the instrumental variables estimator by Angrist, Imbens, and Rubin (1996).[43] It can be calculated from the initial overall average impact estimate and information on which (or how many) intervention group members participate in the program and which do not. It also has the crucial property that it cannot suffer from selection bias due to pre-existing differences between participants and no-shows or participants and comparison group members. Specifically, if the original experimental comparison of average outcomes between all Head Start group members and all non-Head Start group members is not biased by systematic differences between these two randomly generated groups at baseline, the simple rescaling of the original estimate cannot be biased. This theorem, based solely on the assumption of zero impacts on non-participants, provides a broadly accepted basis for the now almost universal practice of reporting impact estimates for participants-only along with the all-intervention-group impact findings.[44]

---

[41] *Impact (average Head Start group member) = Impact (all)/N (all)*

[42] Bloom, H.S. (1984) "Accounting for no-shows in experimental evaluation designs." ***Evaluation Review, 8,*** pp. 225-246.

[43] Angrist, J.D., G.W. Imbens, and D.B. Rubin. (1996). "Identification of causal effects using instrumental variables," ***Journal of the American Statistical Association, 91***, pp. 444-472.

[44] The National Early Head Start Evaluation, for example, reports primarily "no-show-adjusted" estimates of impact on participants rather than highlighting more prominently the more directly obtained impact findings for the average intervention group member.

Adjusting for Head Start Participation by Members of the Non-Head Start Sample

The study had no way to fully ensure that the children and families randomly assigned to the non-Head Start group did not participate in federally funded Head Start. The grantees and delegate agencies whose applicants made up the research sample agreed not to serve those families using Federal Head Start funds during the 2002-03 program year. But other grantees and delegate agencies in nearby communities (or, in the case of several large cities, in overlapping neighborhoods) did not enter into such agreements and, for reasons of privacy, could not be told the identities of the children and families involved in the study even had agreement been reached not to serve them. Moreover, no mechanisms existed for enforcing the commitments made by the participating grantees and delegate agencies.

In light of these limitations and the strong attraction of Head Start to many families, it is not surprising that a number of families from the non-Head Start sample in fact obtained Head Start services for their children during that year. A total of 17.6 percent of the children in the non-Head Start group are known to have participated in a federally funded Head Start program for at least 1 day during the analysis period once analysis weights are applied. Though some of these enrollments may have been very brief, Head Start likely had some of the same impacts in this subset of the comparison group as it did for those randomly assigned to the Head Start sample. If so, measured impacts from the comparison of the two complete samples will **understate** the average impact of the intervention, for all children granted access to Head Start and, following the adjustment described in the previous section, for children in the Head Start sample who actually participated in the program. The consequences of this "contamination" of the research design may be slight and (as discussed earlier in the report) have precedence among randomized evaluations of social programs. Still, it is important to take program participation by the non-Head Start sample into account in looking at the findings and, if possible, to develop measures of impact that reduce or remove any potential "contamination bias" that may have occurred as a result.

A number of strategies have been suggested for analyzing members of a randomly selected comparison group who receive an intervention. These individuals are referred to in the literature as "crossovers" based on the fact that they "crossed over" the line between the two situations created at random assignment. One option for accounting for these individuals (generally viewed as unacceptable) is to assume that the program had no impact on them and interpret unadjusted findings as reflective of the intervention's full impact. This may be a good first approximation, depending on the duration and intensity of program involvement among comparison group members and the number of such individuals. But it is at best a lower bound for the desired measure,[45] Head Start's impact on the average participant compared to a statistically equivalent group with no Head Start participation at all. If no exploration is done of the potential consequences of crossover behavior, the study will be able to say with full confidence that "Head Start's average impact is at least this big, and possibly bigger," but it will not be able to say with confidence what the **upper** extreme might be in terms of Head Start's impact.

Of the three known approaches to addressing the possible magnitude of the canceling-out problem for crossovers (see below), the current report focuses on the most intuitively plausible and straightforward method: removing the "contaminated" cross-over cases from the non-Head Start sample and recalculating the impact of the program without them. In using this strategy, it is acknowledged that the resulting estimates **are no longer fully experimental**, i.e., they do not emerge from the comparison of two complete sets of individuals made statistically equivalent through random assignment. Specifically, if families who obtain access to Head Start despite having been assigned to non-Head Start status differ from other families randomly assigned, the removal of the crossovers will change the composition of the remaining non-Head Start sample, i.e., the latter set of individuals will no longer match the complete Head Start sample to which it is compared. Moreover, the counterparts to crossovers cannot be removed from the Head Start sample; they cannot be identified, since no one knows which children randomly assigned to enter Head Start in the study sites would still have managed to participate in the program had they been assigned originally to the non-Head Start comparison group.

The mismatch caused by removal of crossovers from the analysis sample creates a new form of possible bias, not contamination bias (since the contaminated crossovers have been removed) but selection bias because the removed cases were selected non-randomly. If selection bias exists, it may skew the impact estimates up or down, depending on the ways crossover children differ from other children in the non-Head Start sample. For example, it may be that non-Head Start sample children who gain access to the program face greater developmental challenges than other children and fare worse in the spring regardless of their program

[45] The unadjusted estimate is a lower bound because (a) if true impact on crossovers is zero, it gives the correct answer and (b) if true impact on crossovers is more than zero and in the same direction as the program's impact on participants in the Head Start group, it is too low by virtue of the impact on crossovers canceling out a portion of the impact on participants when the overall non-Head Start and Head Start groups are compared.

participation. This would be the case if the families of the most disadvantaged children, and/or the Head Start providers to which those families apply, press particularly hard to obtain a strong pre-kindergarten experience for those children, and in particular a Head Start experience. If this is the case, removing crossovers from the non-Head Start sample will skew the average outcome of the comparison sample upward and thus lead to an understatement of the program's impact. Alternatively, children with relatively favorable prospects may be the most likely to participate in Head Start as crossovers, possibly because their parents work harder to support their growth in a variety of ways, one of which may be obtaining access to Head Start when assigned to the non-Head Start sample. If this occurs, the non-Head Start sample loses some of its highest achieving children in the spring, causing what remains of that group to have artificially low outcomes compared to the full Head Start sample. In this case, the resulting impact estimates following the removal of crossovers will overstate the impact of the program.

The reliability of adjusted impact measures that omit crossovers from the comparison group will hinge on the researcher's ability to measure and adjust for how crossovers differ from non-crossovers. This means the success of the technique will depend on the ability to model selection into the program within the non-Head Start sample, a challenge facing all attempts to measure social program impacts absent random assignment. In the current report, this problem is handled in the same way that nonresponse is treated in sample surveys: the non-crossover members of the control group—who now constitute a non-experimental comparison group rather than an experimental control group—are weighted back up to represent the complete non-Head Start group adjusting weights in cells defined by available baseline characteristics.

After dropping crossovers from the comparison group sample, the analysis weight adjustments used originally to offset missing data caused by non-response to the spring 2003 collection of follow-up data were recalculated. As explained earlier in the report, children and families omitted from impact calculations due to lack of outcome data are "put back" statistically by increasing the analysis weights (i.e., the degree of influence) of other observations with outcome data that have similar background characteristics. This strategy, applied previously (and separately) to the Head Start and non-Head Start samples, was repeated for just the non-Head Start sample with crossovers treated as additional "non-respondents." This technique adjusts for the absence of potentially contaminated control group members from the now trimmed-back analysis sample by increasing the analysis weights applied to similar non-Head Start sample members who did not cross over and who therefore remain in the analysis. Appendix 1.2 on analysis weights explains in detail this re-weighting process. The regression analysis used to calculate impacts then controls for remaining differences in observed background characteristics between the two samples being compared using covariates as described above.

Just as was true of the original nonresponse adjustments to analysis weights and regression modeling of background characteristics, there is no assurance that the combination of these methodologies effectively or completely offsets the potential bias of using an incomplete comparison group sample. Only the distinctive attributes of the missing non-Head Start sample members captured by the stratification variables used to re-weight the data (in this case principally site and child age; see Appendix 1.2 for details) or the background covariates in the regression equations will be compensated for, leaving considerable potential for remaining differences in attributes to skew the cross-over adjusted results. Additional research in future reports will consider how much of an influence this possible remaining selection bias may have on the magnitude of the estimates by exploring the other two strategies in the literature for addressing crossovers in random assignment impact evaluations:

- Sensitivity analyses of how much larger or smaller outcomes (or impacts) for crossovers would have to be—compared with known values of these quantities for other sample members—for the omission of this group to substantially influence the character of or "story" in the findings;
- Construction of specific alternative scenarios to serve as upper and lower bounds on the size of true impacts, including a scenario sometimes used in the literature that treats crossovers symmetrically with the no-show adjustment described in the previous section. Treating crossovers as has been proposed for no-shows requires much stronger assumptions that generally cannot be justified as the principal response to the crossover problem Specifically, one would have to assert that (a) the average impact of Head Start on crossovers equals that on the corresponding children and families in the Head Start sample even though the former entered the program through a different indirect or surreptitious route, often with different provider agencies, and (b) that average impact does not differ appreciably from Head Start's impact on non-crossover-type sample members. Though not justifiable on their face, a reinterpretation of these assumptions as part of an upper bound scenario makes this approach worth pursuing in future research.

For now, the discussion of findings in subsequent chapters focuses on the information the impact study supplies directly, without recourse to these types of quasi-experimental and simulation analysis methods.

Presentation of Results for Participants and Adjusted for Crossovers

Chapters 5-8 present the average impact of access to Head Start (referred to as "intent to treat" estimates). Related appendices present the average impact of Head Start participation, using the no-show adjustment for those outcomes for which overall average impacts or subgroup/moderator impacts are reported as statistically significant. Overall average impacts, presented in the appendix tables for reference, are divided by the corresponding participation rate to obtain average impacts on participants, using subgroup-specific participation rates where appropriate. Tests of the statistical significance of the participant-only impact findings are identical to those of their full-sample antecedents. Given the maintained assumption that Head Start had no effect on non-participants, a zero (non-zero) impact on the entire Head Start group will occur if, and only if, a zero (non-zero) impact occurs for the average participant. Thus, hypothesis test results for all Head Start group members imply hypothesis test results for participants. One can, therefore, reject the null hypothesis that the average impact on participants is zero whenever the full-sample analysis shows impact to be statistically significant for the broader set of all Head Start group members.

Additionally, the appendix tables that summarize statistically significant impacts include a third set of estimates of effects on participants adjusted to remove the influence of crossovers, i.e., of children assigned to the non-Head Start sample who nonetheless participated in Federal Head Start for at least a day during the 2002-03 program year. These estimates are not attenuated by the potential impacts of Head Start on crossover children as is true of other impact findings, but they may be subject to uncorrected selection bias up or down. Only impacts for an entire age group are examined in this way, not the more detailed findings for subgroups and moderating factors. After computing crossover-adjusted impacts as described earlier in this chapter, the "no-show" adjustment is applied to these estimates to convert results into average impacts on participants for inclusion in the tables. Separate tests of the statistical significance of the crossover-adjusted estimates are presented with the findings, constructed using the procedures described already for the non-crossover-adjusted analyses but applied to the new estimates from the re-weighted data.

Chapter 5: Impact of Head Start on Children's Cognitive Development

Highlights

Head Start was found to have a positive impact on the cognitive development of children in both the 3- and 4-year-old age groups, with children in the Head Start group being more advanced by spring 2003 than non-Head Start children. However, the magnitude of the statistically significant and positive cognitive impact of Head Start, while meaningful in some skill areas, was small to modest, with the magnitude of estimated impacts varying across skill areas. Specifically:

- The largest impacts were found for direct assessments of pre-reading skills (19 to 24 percent of a standard deviation) and parent perceptions of children's emergent literacy skills (effect sizes of 29 to 34 percent of a standard deviation).
- Relatively small impacts were found for the direct assessments of pre-writing skills (13 to 16 percent of a standard deviation) and vocabulary knowledge (effect sizes of 10 to 12 percent of a standard deviation).
- No overall significant impact was found in the skill areas of oral comprehension, phonological awareness, and early math skills. In the latter areas, however, some demographic subgroups showed significant impacts.
- The evidence of beneficial cognitive impacts of Head Start was more widespread and consistent among 3-year-old children than among 4-year-olds.[1]
- Positive cognitive impacts of Head Start were found for children from English-speaking families and to a more limited degree for children from Spanish-speaking families (i.e., children whose home language is Spanish, excluding study children in Puerto Rico). However, tests of the interaction between language and program impact failed to reach statistical significance.
- There was substantial evidence that children in the 3-year-old group whose primary caregivers reported high levels of depressive symptoms at baseline showed less benefit of Head Start on their cognitive development than children whose mothers reported lower or no evidence of depressive symptoms.
- Evidence that Head Start has positive benefits is particularly strong for Hispanic and African American children.

[1] Future analyses will test the statistical significance of the difference in impacts across the two age groups.

Organization and Presentation of Findings

This chapter focuses on the impact of Head Start on six different constructs that make up the cognitive domain:

- Pre-reading skills;
- Pre-writing skills;
- Vocabulary knowledge;
- Oral comprehension and phonological awareness;
- Early math skills; and
- Parent's perceptions of their child's early language and literacy skills.

The discussion that follows focuses on an examination of statistically significant **"intent-to-treat" impact estimates.** This involves using the complete sample of children who were randomly assigned in 2002 (see Chapter 4), measuring the average impact of access to Head Start. The discussion begins with a review of the overall average impacts for all newly entering children in the 3- and 4-year-old groups respectively. Then it examines impacts for subgroups of children defined by the language used for the child assessment (i.e., children who were assessed in English in both fall 2002 and spring 2003, and children who were assessed in Spanish in fall 2002 and in English in spring 2003). Following the review of overall average effects, the discussion examines the extent to which impacts occurred for key subgroups of Head Start children and how different in size impacts may have been for various subgroups.

The estimated **impacts on program participants** (i.e., referred to as the "impact on the treated") are presented in Appendix 5.1, focusing primarily on the extent of any differences from the intent-to-treat estimates. For clarity, the discussion in Appendix 5.1 examines only the combined group of all children (i.e., separate breakdowns by the language of assessment subgroups are not provided).

The statistical results for the discussion in this chapter are presented in a series of tables provided at the end of the chapter, plus additional tables in Appendix 5.2. Exhibits 5.1-A through 5.1-C (for children in the 3-year-old group) and 5.2-A through 5.2-C (for the 4-year-old group), present the overall average impact estimates for the combined sample and for the two separate language groups. The data in these tables are presented for individual measures (e.g., Woodcock-Johnson III Letter-Word Identification) in three ways: (1) as simple mean differences, (2) using regression analyses that include only demographic covariates measured in fall 2002, (3) and using regression analyses that add a measure of the outcome variable assessed in fall 2002.

For the latter two regression-based estimates, the shaded columns in the tables show statistically significant results that are discussed in the text. As discussed in Chapter 4, impact estimates are based on the regression model that includes family demographic characteristics **and** the fall measure of the parent outcome as covariates *only in instances where analyses show the fall measure could not be substantially influenced by early exposure to Head Start.* In those instances where such early substantial influence on the fall measure cannot be ruled out, the regression-adjusted impact estimates use only the family demographic characteristics as covariates. However, because inclusion of the fall outcome measure has other benefits as discussed in Chapter 4 (e.g., increased statistical precision, reduction in potential spring 2003 nonresponse bias) that may be considered more important to some than the removal of some of Head Start's impacts from the spring estimate, the analysis that controls for the fall outcome is provided as well. Conversely, since some risk of a small amount of impact removal exists any time the fall outcome measure enters the analysis, estimates based solely on demographic variables are also presented, although they are not always highlighted in the discussion.

Exhibits 5.3-A and 5.3-B (for the 3- and 4-year-old groups, respectively) summarize all of the statistically significant average impacts both for the overall group and for a set of 11 subgroups identified for examination because of their special program or child development importance as discussed in Chapter 4. Subgroup results are shown first for estimated **differences** in impacts between subgroups (e.g., boys versus girls) and then as **impacts on individual subgroups** (e.g., impacts on boys alone). Both perspectives are important, as discussed in Chapter 4: differences in the size of impacts indicate the types of children and families not benefiting as much as others from Head Start participation; impacts on individual subgroups show where any overall gains from the program are occurring and whether those gains are widespread (as opposed to concentrated among certain segments of the Head Start population). These tables present two columns of figures: the estimated impacts and the estimates expressed as "effect sizes" (i.e., the impact estimates divided by the standard deviation of the outcome measure in the population). Effect sizes provide a yardstick for gauging the quantitative importance of a measured impact in relation to the natural variation of the child or family outcome Head Start is seeking to affect.[2] Effect sizes are important in interpreting the size of Head Start's measured impact and, in particular, how much larger that impact may be for the average program participant as opposed to the larger group of children and families accorded access to the program (some of whom do not participate in the program).

Finally, Exhibits 5.4 through 5.25, provided in Appendix 5.2, show the results of all moderator/subgroup analyses, including those that do not produce statistically significant impacts, with a separate table for each individual measure of cognitive outcomes. Again, for clarity, these results are only presented for the full combined sample (i.e., not separately for the English-English and Spanish-English language groups).

Estimated Impact of Access to Head Start

This first section discusses the estimated impact of Head Start on cognitive outcomes using the sample of children randomly assigned to either Head Start or to the non-Head Start group, referred to as "intent-to-treat" impact estimates. These estimates show the effect of Head Start on the average child given access to the program.

Impact on Pre-Reading Skills

Overall, there were significant impacts on the Pre-Reading Skills of children in both the 3- and 4-year-old groups, with the skills of children in the Head Start group being more advanced than those of non-Head Start group children. Significant differences were found both in children's performances on the Woodcock-Johnson III Letter-Word Identification test and on the Letter Naming task. Both of these tests tap letter recognition skills that are important steps toward becoming a proficient reader and are predictive of how well children are reading at the end of kindergarten and 1st grade.

As shown in Exhibit 5.1-A, among children in the 3-year-old group from all language backgrounds, the IRT scale score on the Woodcock-Johnson III Letter-Word Identification test was 5.65 points higher for the Head Start group than for the non-Head Start group, by the end of the program year. By the same time, the Head Start group children could also identify an average of 1.3 more letters than children in the non-Head Start group (the latter could identify an average of 3.8 letters).

---

[2] The standard deviation of each outcome measure is derived from data on children/families in the non-Head Start sample, excluding members of the Head Start sample, to ensure that any effect of the intervention on the variation of the outcome is excluded from the calculation.

Among children in the 4-year-old group from all language backgrounds (Exhibit 5.2-A), the IRT scale score on the Woodcock-Johnson III Letter-Word Identification test was 5.74 points higher for the Head Start group than for the non-Head Start group. By the same time, the Head Start group children could identify an average of 2.3 more letters than children in the non-Head Start group (the latter could identify an average of 9.2 letters).

The magnitude of the Head Start impact on children's Pre-Reading Skills was modest but meaningful. The effect size of the impacts on Letter-Word Identification test scores were 24 percent of a standard deviation for children in the 3-year-old group and 22 percent for children in the 4-year-old group (see Exhibit 5.3-A and 5.3-B). The effect sizes of the impacts on Letter Naming task performance were 19 percent for children in the 3-year-old group and 24 percent for the 4-year-old group.

The availability of publisher norming data for the Woodcock-Johnson III Letter-Word Identification test permits comparisons of the skill levels of children in the Head Start Impact Study with those of the general population of 3- and 4-year-olds in the US (including those who were not from low-income families). These comparisons showed that, at the end of the program year, the mean performance of Head Start children was still below average performance levels for all U.S. children, by about one-third of a standard deviation. But, by comparing their performance in spring 2003 to the standard scores of children in the non-Head Start group, it appears that Head Start serves to narrow the gap between the skills of Head Start children and the skills of the general population of young children by about 45 percent.

This was determined by estimating the mean standard score on the Woodcock-Johnson III Letter-Word Identification test for children in the 3-year-old Head Start group, which was calculated to be equal to 96.0 in the spring of the program year. The comparable mean standard score for children in the non-Head Start group was estimated to be equal to 92.4. The mean standard score for all U.S. 3-year-olds is 100.0, with a standard deviation of 15. Therefore, the gap between the average score of the 3-year-old Head Start group and the overall national average score was 4 standard-score points; whereas, the gap between the average for the non-Head Start group and the national norm was 7.6 points. Hence, the Head Start group had a deficit that was smaller by 3.6 points, or 47 percent (3.6/7.6 = 0.47). Similarly, for 4-year-olds, the mean standard score on the Letter-Word Identification test for the Head Start group was 95.2 in the spring of the program year; whereas, the mean for 4-year-olds in the non-Head Start group was 91.3. Again, the mean standard score for all U.S. 4-year-olds is 100.0, with a standard deviation of 15. Consequently, the gap between the average score of the 4-year-old Head Start group and the overall national average score was 4.8 standard-score points, and the gap between the average for the non-Head Start group and the national norm was 8.7 points. Hence, the Head Start group had a deficit that was smaller by 3.9 points, or 45 percent (3.9/8.7 = 0.45).

Impact on Pre-Writing Skills

The pre-writing skills of both 3- and 4-year-old children in the Head Start group were slightly more advanced than those of the non-Head Start group children in the same age group. Pre-writing skills were measured by two tests: the McCarthy Draw-A-Design test, which measures perceptual-motor skills involved in seeing and copying basic geometric shapes, and the Woodcock-Johnson III Spelling test, which first measures perceptual-motor skills involved in tracing or copying letter shapes and then measures children's ability to draw letters on request, without being shown the shape of the letter in question. Significant differences were found both in children's performances on the Woodcock-Johnson III Spelling test and on the McCarthy Draw-A-Design test. Both of these tests tap skills that are necessary for writing words and sentences and are also predictive of how well children are reading at the end of kindergarten and 1st grade.

Among children in the 3-year-old group, a statistically significant difference in average scores was found on the Draw-A-Design test between the Head Start group and the non-Head Start group children. However, no significant difference was found for children in the 3-year-old group on the (more advanced) Spelling test. On the other hand, for children in the 4-year-old group, a significant difference in average scores was found on the Spelling test between children in the Head Start and non-Head Start groups, but no statistically significant difference was found for these older children on the (more basic) Draw-A-Design test.

As shown in Exhibit 5.1-A, among children in the 3-year-old group from all language backgrounds, the score on the McCarthy Draw-A-Design test was 0.15 points higher for the Head Start group than for the non-Head Start group. However, the IRT scale score on the Woodcock-Johnson III Spelling test for children in the 3-year-old Head Start group was not significantly different from that of children in the non-Head Start group. For children from all language backgrounds in the 4-year-old group (see Exhibit 5.2-A), the IRT scale score on the Woodcock-Johnson III Spelling test was 4.14 points higher for the Head Start group than for the non-Head Start group. Mean scores of the two groups of children in the 4-year-old group on the McCarthy Draw-A-Design test were not significantly different.

The magnitude of the Head Start impact on children's Pre-Writing Skills, though statistically significant, was relatively small. The effect size of the impact on McCarthy Draw-A-Design test scores of children in the 3-year-old group was 13 percent, and the effect size of the impact on the Woodcock-Johnson III Spelling test scores of children in the 4-year-old group was 16 percent of a standard deviation.

As above, the availability of publisher norming data for the Woodcock-Johnson III Spelling test allows a comparison of the skill levels of children in the Head Start Impact Study with those of the general population of 3- and 4-year-olds in the US (including those who were not from low-income families). These comparisons showed that, at the end of the program year, the mean performance of Head Start 4-year-old children was still below average performance levels for all U.S. 4-year-olds by one-half of a standard deviation. At the same time, based on the mean standard scores of children in the non-Head Start group, it would appear that Head Start narrows the gap between the early writing skills of Head Start-eligible children and the skills of the general population of young children by 28 percent.[3]

Impacts on Vocabulary Knowledge

Among children in the 3-year-old group, the Vocabulary Knowledge of children in the Head Start group was slightly more advanced than that of the non-Head Start children in the same age group. Among children in the 4-year-old group, only Head Start children from Spanish-speaking families showed vocabulary knowledge significantly greater than that of the non-Head Start children. Vocabulary knowledge was measured by two tests: the Peabody Picture Vocabulary Test, Third Edition (PPVT-III) (adapted), which measures children's receptive vocabulary by asking them to select one of four pictures that best represents each of a series of words spoken by the examiner, and a Color Naming test that measures children's ability to name the colors of drawings of bears in 10 different colors. Both of these tests tap skills that are indicative of children's oral language development. Vocabulary tests are strongly predictive of children's general knowledge at the end of kindergarten and 1st grade. Vocabulary is also predictive of later reading proficiency as children move from the basic decoding stages of reading

[3] The mean standard score on the Spelling test for children in the 4-year-old group randomly assigned to Head Start was 92.5 in the spring of the program year, whereas the mean for children in the non-Head Start group was 89.6. (The mean standard score for all U.S. 4-year-olds is 100.0, with a standard deviation of 15.) The gap between the average score of the 4-year-old Head Start group and the overall national average score was 7.5 standard-score points; whereas, the gap between the average for the non-Head Start group and the national norm was 10.4 points. Hence, the Head Start group had a deficit that was smaller by 2.9 points, or 28 percent (2.9/10.4 = .28).

to more advanced literacy that involves relating what is read to information already acquired about the outside world.

As shown in Exhibit 5.1-A, among children in the 3-year-old group from all language backgrounds, the IRT scale score on the PPVT-III (adapted) was 4.23 points higher for the Head Start group than for the non-Head Start group, and children in the Head Start group could name about one more color than children in the non-Head Start group (the estimated impact is 0.70). As shown in Exhibit 5.2-A, among children in the 4-year-old group from all home-language backgrounds, the IRT scale score on the PPVT-III (adapted) was not significantly different from that of children in the non-Head Start group, nor was there an overall average impact found for these children on the Color Naming test. The magnitude of the Head Start impact on children's Vocabulary Knowledge, though statistically significant, was relatively small. The effect size of the impact on PPVT-III (adapted) scores of children in the 3-year-old group was 12 percent of a standard deviation and about 10 percent for the Color Naming task.

Again, using publisher norming data for the PPVT-III shows that Head Start children were 8 percent[4] closer to the national norm on vocabulary knowledge than non-Head Start children, but only for children in the 3-year-old group.

Impacts on Oral Comprehension and Phonological Awareness

Two other skill areas that have been shown to relate to children's emergent literacy and later academic achievement are oral comprehension and phonological awareness. Oral comprehension is the child's ability to understand and make inferences from spoken phrases and sentences. In the Head Start Impact Study, this skill was measured by the Woodcock-Johnson III Oral Comprehension test. The test consists of a series of incomplete sentences spoken to the child. The child is asked to "fill in the blank" in each sentence based on contextual cues contained in the sentence and his or her prior knowledge of common phrases.

Phonological awareness is a child's understanding that spoken sentences are made of component words, compound words are made up of simpler words, and that even simple words

[4] The mean standard score on the PPVT-III for children in the 3-year-old group randomly assigned to Head Start was 82.9 in the spring of the program year and 81.4 for children in the non-Head Start group. (The mean standard score for all U.S. 3-year-olds is 100.0, with a standard deviation of 15.) The gap between the average score of the 3-year-old Head Start group and the overall national average score was 17.1 standard-score points; whereas, the gap between the average for the control group and the national norm was 18.6 points. Hence, the Head Start group had a deficit that was smaller by 1.5 points, or 8 percent (1.5/18.6 = .08).

are made up of component syllables and sounds (phonemes). The skill also involves understanding that when a component sound is added to or deleted from a word, the meaning of the resulting word is often different from that of its unaltered predecessor. In the Head Start Impact Study, phonological awareness was measured by the Elision task from the Comprehensive Test of Print and Phonological Processing – Preschool edition (CTOPPP). In this task, children were asked to identify the word that resulted when part of a word was deleted. The test first uses pictures to assist children in determining the right answer, but then progresses to asking about the results of word, syllable, or phoneme deletion without any picture assistance.

Neither children's scores on the Woodcock-Johnson III Oral Comprehension test nor their scores on the CTOPPP Elision test showed any significant overall effect of access to Head Start (see Exhibits 5.1 and 5.2), i.e., the scores of children in the Head Start group did not differ significantly at the end of the program year from those of children in the non-Head Start group. This was the case among children both in the 3-year-old and 4 year-old groups, in all language groups combined, as well as among children from English-speaking families and Spanish-speaking families analyzed separately.

Impact on Early Math Skills

For the most part, the Early Math Skills of 3- and 4-year-old children did not show a significant impact resulting from access to Head Start. There were, however, some analytic subgroups that showed significant positive impacts of Head Start on early math achievement.

Math skills were measured with two tests: the Woodcock-Johnson III Applied Problems test and a rating scale that assessors used to evaluate how well children had done at one-to-one counting of a set of drawings of 10 bears (the same ones used in the Color Naming task). The Applied Problems test assesses children's proficiency at solving simple word problems that involve counting, simple arithmetic, and basic measurement. Both of these tests assess basic skills and understandings that are essential for the development of more advanced quantitative capabilities and are predictive of mathematics achievement in kindergarten and 1st grade. The test scores are also predictive of later reading achievement because they rely on children's comprehension of spoken questions and instructions and on children's ability to make one-to-one associations between sounds and pictures or written symbols.

Although the mean scores on the Applied Problems test of children in both the 3- and 4-year-old Head Start groups were generally higher than those of non-Head Start children, these differences did not reach statistical significance. Differences in the ratings of how well children did on one-to-one Counting Bears were also generally small and not statistically reliable.

Impact on Parent Perceptions of Children's Emerging Literacy Skills

Among children in both age groups, parental reports of children's emergent literacy skills were higher for children in the Head Start group compared to those in the non-Head Start group. Parent perceptions of their children's early academic skills were measured by the Parent-Reported Emergent Literacy Scale (PELS). This is a series of questions about how many letters of the alphabet the child knows, how many colors he or she can identify, how high he or she can count, whether the child can write his or her first name, etc. These questions were first developed for the 1993 National Household Education Survey on School Readiness, and resulting scale scores have been shown to correlate with children's age and disability status, with socioeconomic family characteristics, and with other measures of children's cognitive and social development.

As shown in Exhibit 5.1-A, among children in the 3-year-old group from all language backgrounds, the PELS scale score for children in the Head Start group was about 0.5 points higher than the scores for children in the non-Head Start group. Among children in the 4-year-old group from all language backgrounds (Exhibit 5.2-A), the PELS score for children in the Head Start group was 0.4 points higher. The magnitude of the Head Start impact on parent perceptions of children's emerging literacy skills was moderate, with effect sizes of 34 percent of a standard deviation for 3-year-old group children and 29 percent of a standard deviation for 4-year-old group children.

Moderator/Subgroup Differences

The analysis of impacts by subgroups of children and families (detailed in Appendix 5.2 and summarized in Exhibits 5.3-A and 5.3-B for those found to be statistically significant) show some variations in impact for particular types of Head Start participants. The most important and consistent of these findings are discussed below in two sections. The first section looks at instances where a statistically significant **difference** in impact was found between particular subsets of children that were identified in advance as being of program or child development importance, e.g., larger or smaller impacts for special needs children as compared to children without special needs. The second section looks at statistically significant impacts on particular subsets of children, e.g., the impact on children with special needs. The former tells us whether one type of Head Start child or family is benefiting more than another, while the latter points out where impacts are occurring and whether they are widespread. The former "difference-in-differences" impacts are more difficult to detect (with a given sample size); relatively few such impact variations have been detected.

Differences in Impact

The subgroup factors will at times lead to discernibly **different** sizes of impact for one subpopulation than another, information that may be helpful for assessing and enhancing the program. In the cognitive domain, a statistically significant relationship was found between the primary caregiver's reported level of depressive symptoms and the impact of Head Start across several of the cognitive outcome measures, but only for children in the 3-year-old group. Consistently, the impact of Head Start was found to **decrease** with increasing levels of caregiver's reported depressive symptoms (see Exhibits 5.3-A and 5.3-B). The areas of development in which such interactions were found were vocabulary knowledge (on both PPVT-III adapted and Color Naming), early math skills (both Woodcock-Johnson III Applied Problems and Counting Bears), phonological awareness (CTOPPP Elision), and parent perceptions of children's emerging literacy (the PELS measure).

Other statistically significant findings in Exhibits 5.3-A and 5.3-B are not discussed because it is possible they are due to chance alone and do not represent true impacts of the intervention (see discussion of subgroup impact analysis in Chapter 4).[5]

Impacts on Particular Subgroups

Child Language. This section summarizes information from both the language-group-specific analyses (Exhibits 5.1 and 5.2), and the analysis of subgroup impacts defined by language using the combined sample (Exhibits 5.3-A and 5.3-B). As shown in these tables, significant impacts of Head Start were found for English-speaking children in the areas of pre-reading, pre-writing, and vocabulary skills, but only in the area of vocabulary for Spanish-speaking children.[6]

---

[5] While each of the remaining subgroup findings taken one at a time is structured to limit the probability of a "false positive" to 1 in 20, as a group it is almost inevitable that some of these results will reach that level by chance alone. Only when a substantial share of all the tests of impact conducted for a given subgroup—or of a difference in impact between two subgroups—is statistically significant across all four of the outcome domains considered (not simply the outcomes reported in this chapter) can we be sure that at least some of those findings represent real impacts.

[6] As discussed elsewhere, this excludes children from Puerto Rico.

With regard to **pre-reading**, English-speaking children in both age groups exhibited positive impacts on the Woodcock-Johnson III Letter-Word Identification test, and children in the 4-year-old group also had positive impacts on the Letter naming task. In the area of **pre-writing** skills, English-speaking children in the 3-year-old group had positive impacts on the easier Draw-a-Design test, while children in the 4-year-old group exhibited a positive impact on the more difficult Woodcock-Johnson III Spelling test (this same age group difference was also found in the overall average impacts). Finally, in terms of **vocabulary,** significant impacts were found for 3 year-old Head Start group children from English-speaking families (on the Color Naming task only) and from Spanish-speaking families (on both the PPVT-III (adapted) and Color Naming). As shown in Exhibit 5.1-C, among children in the 3-year-old group from Spanish-speaking families, the IRT scale score on the PPVT-III (adapted) was about nine points higher for the Head Start group than for the non-Head Start group. The score on the Color Naming test for these children was 2.52 points higher than that of children in the non-Head Start group, meaning that the Spanish-speaking Head Start children could recognize more than two additional colors than the non-Head Start group of Spanish-speaking children. For these Spanish-speaking children, these gains represent a reduction in their deficit from national norms on the PPVT-III of 13 percent[7].

Head Start also had an impact on parent perceptions of children's emerging literacy (the PELS measure), for both English- and Spanish-speaking families in the 3-year-old group. For the 4-year-old group, an impact was only found for the English-speaking families.

Race and Ethnicity. The results for race and ethnicity are somewhat scattered across the cognitive domains, but there is particularly strong evidence that Head Start is having a positive impact on the cognitive development of minority children. Specifically, for Hispanic children in the 3-year-old group, positive impacts are noted in pre-reading (both Woodcock-Johnson III Letter Word Identification and Letter Naming), vocabulary (PPVT-III, adapted), and pre-writing (Woodcock-Johnson III Spelling). For African American children in the 3-year-old group, positive impacts are noted in pre-reading (Woodcock-Johnson III Letter Word Identification), phonological awareness (CTOPPP Elision), and in pre-writing (Draw-a-Design); for African American children in the 4-year-old group, positive impacts are noted in pre-reading skills

[7] The gap between the average score of children in the 3-year-old Head Start subgroup from Spanish-speaking families and the overall national average score was 24.7 standard-score points, whereas the gap between the average for the non-Head Start group and the national norm was 28.3 points. Hence, the Head Start group had a deficit that was smaller by 3.6 points, or 13 percent (3.63/28.3 = .13).

(Woodcock-Johnson III Letter-Word Identification), and on early writing (Woodcock-Johnson III Spelling).

Additionally, for White children in the 3-year-old group, positive impacts are noted in oral comprehension (Woodcock-Johnson III Oral Comprehension); for White children in the 4-year-old group, positive impacts are noted in pre-reading skills (Letter Naming Task).

Head Start also had an impact on parent perceptions of children's emerging literacy (the PELS measure), for African American, Hispanic, and White children in the 3-year-old group. For the 4-year-old group an impact was only found for African-American children.

Parent's Marital Status. Although impacts for these subgroups are noted on several cognitive measures, the estimated impacts are scattered across domains and inconsistent in terms of whether greater impacts are noted for children with married parents or for children with unmarried parents. Consequently, at this time, no conclusions can be drawn from these results, and further analysis is needed.

Special Needs Children. Impacts are found in both pre-reading and pre-writing skills, but in all cases, the impact of Head Start is found for children **without** special needs.

In total, the subgroup-specific findings indicate that cognitive gains from Head Start participation are widespread across the different demographic groups examined, with all groups except special needs children shown to be benefiting.

Exhibit 5.1-A: Initial One-Year Estimates of the Impact of Access to Head Start on Cognitive Outcomes: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group (Weighted Data)

Outcome Measure	**Intent-To-Treat Impact Estimates**				
	Head Start Mean	**Non-Head Start Mean**	**Mean Difference[2]**	**Regression-Adjusted Impact Estimates (Demographic Covariates Only)**	**Regression-Adjusted Impact Estimates (With Fall Measure)[3]**
(Sample N= 2,071):					
PPVT—IRT/ML	254.0	250.0	4.0*	3.38	4.23* (0.12)
CTOPPP Elision (English)	243.4	239.7	3.6	3.51	4.10
Woodcock-Johnson: Letter-Word Identification—IRT/ML[1]	307.0	300.5	6.5***	5.65** (0.24)	5.53***
Letter Naming Task	5.5	3.9	1.6**	1.22*	1.30* (0.19)
Woodcock-Johnson: Spelling—IRT/ML	346.6	343.6	2.9	2.05	2.40
Woodcock-Johnson: Applied Problems—IRT/ML	377.3	373.6	3.7	3.62	4.04
Woodcock-Johnson: Oral Comprehension—IRT/ML	435.5	435.4	0.1	0.35	0.62
Color Naming/Identification	13.9	13.0	0.9*	0.87	0.70* (0.10)
McCarthy Drawing	3.2	3.0	0.2**	0.14*	0.15* (0.13)
Counting Bears	2.9	2.7	0.2	0.14	0.12
Parent Educational Literacy Activities Scale (PELS)[1]	2.9	2.4	0.5***	0.47 *** (0.34)	0.42***

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.
[1] Fall measure used in regression failed statistical test.
[2] Differences are rounded to the nearest 0.1.
[3] As described in Chapter 4, two regression specifications were estimated for some of the cognitive outcomes in the combined sample. The two models yielded the same results.

Note: Numbers in parentheses in shaded boxes are estimated effect sizes.

Exhibit 5.1-B: Initial One-Year Estimates of the Impact of Access to Head Start on Cognitive Outcomes: 3-Year-Old Group, Fall English-Spring English Group (Weighted Data)

Outcome Measure	Intent-To-Treat Impact Estimates				
	Head Start Mean	Non-Head Start Mean	Mean Difference[2]	Regression-Adjusted Impact Estimates (Demographic Covariates Only)	Regression-Adjusted Impact Estimates (With Fall Measure)
(Sample N= 1,629):					
PPVT—IRT/ML	259.0	255.5	3.5	2.85	3.70
CTOPPP Elision (English)	248.2	242.9	5.3	4.70	3.89
Woodcock-Johnson: Letter-Word Identification—IRT/ML	308.8	302.7	6.1**	5.43**	5.19*** (0.15)
Letter Naming Task	5.7	4.3	1.4*	1.13	1.22* (0.18)
Woodcock-Johnson: Spelling—IRT/ML	346.1	344.4	1.7	1.12	0.95
Woodcock-Johnson: Applied Problems—IRT/ML[1]	381.3	378.2	3.0	3.04	2.94
Woodcock-Johnson: Oral Comprehension—IRT/ML	438.9	438.4	0.5	0.43	0.22
Color Naming/Identification	14.7	14.1	0.7	0.66	0.56
McCarthy Drawing	3.2	3.0	0.2*	0.14*	0.15* (0.13)
Counting Bears	2.9	2.8	0.1	0.12	0.10
Parent Educational Literacy Activities Scale (PELS)	3.0	2.5	0.5***	0.48***	0.44*** (0.19)

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.
[1] Fall measure used in regression failed statistical test.
[2] Differences are rounded to the nearest 0.1.

Note: Numbers in parentheses in shaded boxes are estimated effect sizes.

Exhibit 5.1-C: Initial One-Year Estimates of the Impact of Access to Head Start on Cognitive Outcomes: 3-Year-Old Group, Fall Spanish-Spring English Group (Weighted Data)

Outcome Measure	Intent-To-Treat Impact Estimates				
	Head Start Mean	**Non-Head Start Mean**	**Mean Difference[2]**	**Regression-Adjusted Impact Estimates (Demographic Covariates Only)**	**Regression-Adjusted Impact Estimates (With Fall Measure)**
(Sample N= 442):					
PPVT—IRT/ML[1]	234.2	225.1	9.1*	9.09* (0.26)	6.31*
TVIP—IRT /ML[1]	253.4	247.1	6.3	3.24	4.60
CTOPPP Elision (English)[1]	224.0	224.6	-0.5	0.56	0.80
Woodcock-Johnson: Letter-Word Identification—IRT/ML[1]	300.1	291.3	8.8*	8.75* (0.36)	6.88
Woodcock-Muñoz: Letter-Word Identification—IRT/ML[1]	352.2	350.2	2.0	0.64	-0.25
Letter Naming Task[1]	4.6	2.2	2.3*	1.77* (0.27)	1.53
Woodcock-Johnson: Spelling—IRT/ML[1]	348.3	340.3	8.0*	6.79	4.83
Woodcock-Johnson: Applied Problems—IRT/ML[1]	361.7	353.4	8.2	8.68	5.91
Woodcock-Johnson: Oral Comprehension—IRT/ML[1]	422.0	421.3	0.7	1.14	1.08
Color Naming/Identification[1]	10.9	8.7	2.2	2.52* (0.35)	1.77*
McCarthy Drawing[1]	3.3	3.1	0.2	0.18	0.16
Counting Bears[1]	2.5	2.2	0.3	0.30	0.25
Parent Educational Literacy Activities Scale (PELS)[1]	2.4	1.9	0.6**	0.52** (0.22)	0.42***

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.
[1] Fall measure used in regression failed statistical test.
[2] Differences are rounded to the nearest 0.1.

Note: Numbers in parentheses in shaded boxes are estimated effect sizes.

Exhibit 5.2-A: Initial One-Year Estimates of the Impact of Access to Head Start on Cognitive Outcomes: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group (Weighted Data)

Outcome Measure	Intent-To-Treat Impact Estimates				
	Head Start Mean	Non-Head Start Mean	Mean Difference[2]	Regression-Adjusted Impact Estimates (Demographic Covariates Only)	Regression-Adjusted Impact Estimates (With Fall Measure)[3]
(Sample N=1,638):					
PPVT—IRT/ML	293.9	291.3	2.5	2.00	2.59
CTOPPP Elision (English)	275.2	273.7	1.6	1.01	1.40
Woodcock-Johnson: Letter-Word Identification—IRT/ML[1]	325.5	319.2	6.2*	5.74* (0.22)	3.48
Letter Naming Task	11.5	9.2	2.3**	2.23**	2.28** (0.24)
Woodcock-Johnson: Spelling—IRT/ML	371.6	367.7	3.9*	4.01*	4.14* (0.16)
Woodcock-Johnson: Applied Problems—IRT/ML	397.5	394.4	3.0	2.69	2.88
Woodcock-Johnson: Oral Comprehension—IRT/ML	443.4	443.7	-0.2	-1.05	-0.90
Color Naming/Identification[1]	17.1	16.5	0.7	0.60	0.18
McCarthy Drawing[1]	4.5	4.4	0.2	0.22	0.13
Counting Bears	3.8	3.6	0.2	0.16	0.12
Parent Educational Literacy Activities Scale (PELS)[1]	3.8	3.3	0.4***	0.41*** (0.29)	0.29**

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.
[1] Fall measure used in regression failed statistical test.
[2] Differences are rounded to the nearest 0.1.
[3] As described in Chapter 4, two regression specifications were estimated for some of the cognitive outcomes in the combined sample. The two models yielded the same results.

Note: Numbers in parentheses in shaded boxes are estimated effect sizes.

Exhibit 5.2-B: Initial One-Year Estimates of the Impact of Access to Head Start on Cognitive Outcomes: 4-Year-Old Group, Fall English-Spring English Group (Weighted Data)

Outcome Measure	Intent-To-Treat Impact Estimates				
	Head Start Mean	**Non-Head Start Mean**	**Mean Difference[2]**	**Regression-Adjusted Impact Estimates (Demographic Covariates Only)**	**Regression-Adjusted Impact Estimates (With Fall Measure)**
(Sample N=1,130):					
PPVT—IRT/ML	304.3	303.6	0.8	0.85	1.25
CTOPPP Elision (English)	284.9	282.8	2.2	1.85	0.15
Woodcock-Johnson: Letter-Word Identification—IRT/ML[1]	330.2	322.5	7.8*	6.86* (0.26)	4.01
Letter Naming Task	13.1	10.1	3.1**	2.88**	2.99** (0.32)
Woodcock-Johnson: Spelling—IRT/ML[1]	371.8	367.2	4.6*	4.07* (0.16)	3.14
Woodcock-Johnson: Applied Problems—IRT/ML	402.5	400.8	1.7	1.56	1.46
Woodcock-Johnson: Oral Comprehension—IRT/ML	450.1	450.6	-0.5	-0.38	-0.57
Color Naming/Identification[1]	17.9	17.5	0.4	0.44	0.01
McCarthy Drawing[1]	4.5	4.3	0.2	0.26	0.14
Counting Bears[1]	3.8	3.8	0.0	0.02	-0.02
Parent Educational Literacy Activities Scale (PELS)[1]	4.0	3.5	0.5***	0.44*** (0.13)	0.23*

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.
[1] Fall measure used in regression failed statistical test.
[2] Differences are rounded to the nearest 0.1.

Note: Numbers in parentheses in shaded boxes are estimated effect sizes.

Exhibit 5.2-C: Initial One-Year Estimates of the Impact of Access to Head Start on Cognitive Outcomes: 4-Year-Old Group, Fall Spanish-Spring English Group (Weighted Data)

Outcome Measure	Intent-To-Treat Impact Estimates				
	Head Start Mean	**Non-Head Start Mean**	**Mean Difference[2]**	**Regression-Adjusted Impact Estimates (Demographic Covariates Only)**	**Regression-Adjusted Impact Estimates (With Fall Measure)**
(Sample N=508):					
PPVT—IRT/ML [1]	266.9	262.5	4.4	5.92* (0.15)	8.49*
TVIP—IRT/ML [1]	296.0	291.9	4.1	5.93	7.95
CTOPPP Elision (English) [1]	250.4	251.7	-1.3	-0.39	0.67
Woodcock-Johnson: Letter-Word Identification—IRT/ML[1]	313.3	311.7	1.5	1.70	7.31*
Woodcock-Muñoz: Letter-Word Identification—IRT/ML [1]	358.0	357.1	0.9	1.31	2.76**
Letter Naming Task [1]	7.5	7.3	0.1	0.32	1.27
Woodcock-Johnson: Spelling—IRT/ML [1]	371.1	368.8	2.3	2.29	3.90
Woodcock-Johnson: Applied Problems—IRT/ML [1]	384.5	379.9	4.7	5.72	8.29*
Woodcock-Johnson: Oral Comprehension—IRT/ML [1]	425.2	426.6	-1.4	-1.43	-0.73
Color Naming/Identification [1]	15.1	14.2	0.9	0.98	0.61
McCarthy Drawing [1]	4.7	4.6	0.0	0.14	0.12
Counting Bears [1]	3.6	3.2	0.4	0.41	0.41
Parent Educational Literacy Activities Scale (PELS) [1]	3.2	2.9	0.3	0.32	0.29**

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.
[1] Fall measure used in regression failed statistical test.
[2] Differences are rounded to the nearest 0.1.

Note: Numbers in parentheses in shaded boxes are estimated effect sizes.

Exhibit 5.3-A: Initial Estimates of the Impact of Head Start on Cognitive Outcomes, Statistically Significant Results Only, 3-Year-Old Group, Combined English-English and Spanish-English Group (Weighted Data)

Outcome Measures	Estimated Impact of Access to Head Start	Effect Size
Overall Impact		
PPVT-III	4.23*	0.12
WJ-III Letter-Word Identification--IRT/ML	5.65**	0.24
Letter Naming Task	1.30*	0.19
Color Naming/Identification	0.70*	0.10
McCarthy Drawing	0.15*	0.13
PELS	0.47***	0.34
Difference in Impact [1]		
PPVT-III: Depression	-0.11*	- 0.00
CTOPPP Elision: Depression	-0.13*	- 0.00
WJ-III Applied Problems: Depression	-0.09*	- 0.00
WJ-III Oral Comprehension: Race (White Impact Exceeds African American)	4.73*	0.33
Color Naming/Identification: Depression	-0.02*	- 0.00
Counting Bears: Depression	-0.00*	- 0.00
PELS: Depression	-0.00*	- 0.00
Impact on Subgroup [2]		
PPVT-III: Parent Married	4.51*	0.13
PPVT-III: Spanish-English Language Group	7.52*	0.21
PPVT-III: Hispanic	7.26*	0.21
CTOPPP Elision: African American	7.47*	0.17
WJ-III Letter-Word Identification: No Special Needs	5.38**	0.22
WJ-III Letter-Word Identification: African American	5.80**	0.24
WJ-III Letter-Word Identification: Hispanic	6.92*	0.29
WJ-III Letter-Word Identification: English-English Language Group	5.05***	0.21
WJ-III Letter-Word Identification: Parent Married	6.53**	0.27
WJ-III Letter-Word Identification: Parent Not Married	5.21**	0.22
WJ-III Spelling: Hispanic	5.61*	0.25
Letter Naming Task: No Special Needs	1.24*	0.19
Letter Naming Task: Hispanic	1.45*	0.22
Letter Naming Task: Parent Not Married	1.46*	0.22
WJ-III Oral Comprehension: White	2.82**	0.20

Exhibit 5.3-A: (continued)

Outcome Measures	**Estimated Impact of Access to Head Start**	**Effect Size**
WJ-III Oral Comprehension: Parent Married	2.09*	0.15
Color Identification: English-English Language Group	0.87**	0.12
Color Identification: Parent Married	1.50**	0.21
McCarthy Drawing: No Special Needs	0.16*	0.14
McCarthy Drawing: African American	0.18*	0.16
McCarthy Drawing: English-English Language Group	0.17*	0.15
PELS: No Special Needs	0.50***	0.36
PELS: White	0.37***	0.27
PELS: African American	0.53**	0.38
PELS: Hispanic	0.51**	0.37
PELS: English-English Language Group	0.48***	0.35
PELS: Spanish-English Language Group	0.46*	0.33
PELS: Parent Married	0.52***	0.38
PELS: Parent Not Married	0.43***	0.31

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.

[1] A total of 82 differences in impacts between subgroups were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 5.2. Findings for depression indicate the change in Head Start's estimated impact that accompanies a 1-point increase in mother's baseline depression score. Findings for baseline factors other than depression indicate the amount by which Head Start's estimated impact for the first subset of participants listed in the row label exceeds that for the second subset listed.

[2] A total of 99 subgroup impacts were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 5.2.

Exhibit 5.3-B: Initial Estimates of the Impact of Head Start on Cognitive Outcomes, Statistically Significant Results Only, 4-Year-Old Group, Combined English-English Spanish-English Group (Weighted Data)

Outcome Measures	**Estimated Impact of Access to Head Start**	**Effect Size**
Overall Impact		
WJ-III Letter-Word Identification	5.74*	0.22
Letter Naming Task	2.28**	0.24
WJ-III-Spelling	4.14*	0.16
Color Identification	0.60	0.11
PELS	0.41***	0.29
Difference in Impact [1]		
Counting Bears: Race (Hispanic Impact Exceeds White)	0.52*	0.38
Impact on Subgroup [2]		
PPVT-III: Hispanic	5.64*	0.14
WJ-III Letter-Word Identification: No Special Needs	5.88*	0.22
WJ-III Letter-Word Identification: African American	10.56*	0.40
WJ-III Letter-Word Identification: English-English Language Group	7.32*	0.27
WJ-III Letter-Word Identification: Parent Not Married	7.92*	0.30
Letter Naming Task: No Special Needs	2.39**	0.25
Letter Naming Task: White	2.77**	0.29
Letter Naming Task: English-English Language Group	3.05**	0.32
Letter Naming Task: Parent Not Married	2.70*	0.29
McCarthy Drawing: Parent Not Married	0.39*	0.20
WJ-III Spelling: No Special Needs	4.97**	0.19
WJ-III Spelling: African American	9.75**	0.38
WJ-IIII Spelling: English-English Language Group	4.49**	0.17
WJ-III Spelling: Parent Not Married	6.31*	0.25
PELS: No Special Needs	0.43***	0.30
PELS: African American	0.75**	0.53
PELS: English-English Language Group	0.45***	0.32
PELS: Parent Married	0.35*	0.25
PELS: Parent Not Married	0.52**	0.37

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

[1] A total of 82 differences in impacts between subgroups were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 5.2. Findings for depression indicate the change in Head Start's estimated impact that accompanies a 1-point increase in mother's baseline depression score. Findings for baseline factors other than depression indicate the amount by which Head Start's estimated impact for the first subset of participants listed in the row label exceeds that for the second subset listed.

[2] A total of 99 subgroup impacts were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 5.2

Chapter 6: Impact of Head Start on Children's Social-Emotional Development

Highlights

Among children in the 3-year-old group, those in the Head Start group were reported by their parents to show less frequent and severe problem behavior than non-Head Start children in the same age group. However, the magnitude of the positive impact of Head Start on children's social and emotional development, while statistically significant, was relatively small. No overall statistically significant impacts for social-emotional outcomes were found for children in the 4-year-old group overall, but impacts were found for the subgroup of children from English-speaking families. Specifically:

- Positive impacts of Head Start were found on the Total Problem Behavior measure for children in the 3-year-old group (effect size of 13 percent of a standard deviation) and the Hyperactive Behavior measure for the same age group (effect size of 18 percent of a standard deviation).
- A positive impact was also found on the Aggressive Behavior measure for children in the 4-year-old group, but only for those children from English-speaking families (effect size of 15 percent of a standard deviation).
- No overall impact of Head Start was found on parent-reported Social Skills and Positive Approaches to Learning of children, nor on parent-reported Social Competencies, for children in both age groups.

It is important to note that the analysis of Head Start's impact on children's social and emotional development has thus far relied solely on behavior reports from parents. An important additional source of information on children's social development and positive and negative behaviors, reports from children's teachers and caregivers, could not be used at this stage of the impact analysis. This is because such reports on non-Head Start children who only had parent care are not available. Teacher reports on the social behavior of all study children will be available in future years of the study, when the children are in elementary school.

Organization and Presentation of Findings

This chapter focuses on the impact of Head Start on the following three constructs of children's social-emotional development:

- Social skills and approaches to learning;
- Social competencies; and
- Problem behavior (total problem behavior, aggressive, hyperactive, and withdrawn).

As in Chapter 5, the discussion of estimated impacts focuses on an examination of statistically significant **"intent-to-treat" impact estimates** using the complete sample of children who were randomly assigned in 2002. The discussion begins with the overall average impacts for all newly entering children in both the 3- and 4-year-old groups, and then examines any notable differences in average impacts by the language used for child assessment. Finally, the discussion moves to an examination of the extent to which impacts occurred for key subgroups of Head Start children, and how different in size impacts may be for various subgroups. As in Chapter 5, estimated **impact of Head Start on program participants** is provided in Appendix 6.1, examining the combined group of all children (i.e., separate breakdowns by the language of assessment are not provided).

The statistical results for the discussion in this chapter are presented in a series of tables appearing at the end of this chapter and in Appendix 6.2. Exhibits 6.1-A through 6.1-C (for children in the 3-year-old group) and 6.2-A through 6.2-C (for the 4-year-old group), present the overall average impact estimates for the combined sample and for the two separate language groups. As in Chapter 5, the data are presented in three ways: (1) as simple mean differences, (2) using regression analyses that include only demographic covariates measured in fall 2002, (3) and using regression analyses that added a measure of the outcome variable assessed in fall 2002. Shading is used to indicate which estimate is discussed in the text. Also similar to Chapter 5, Exhibits 6.3-A and 6.3-B (for the 3- and 4-year-old groups, respectively) summarize all of the statistically significant impacts (both for the overall group and for a set of 15 subgroups discussed in Chapter 4) and provide the average impact of access to Head Start along with associated effect sizes. Finally, Exhibits 6.4 through 6.15, provided in Appendix 6.2, show the results of the moderator/subgroup analyses, with a separate table for each individual measure of social-emotional outcomes (again, only for the full combined sample).

Estimated Impact of Access to Head Start

This first section discusses the estimated impact of Head Start on social-emotional outcomes using the sample of children randomly assigned to either Head Start or to the non-Head Start group, referred to as "intent-to-treat" impact estimates. These estimates show the effect of Head Start on the average child given access to the program.

Impact on Social Skills and Approaches to Learning

As reported by parents, the Social Skills and Positive Approaches to Learning (SSPAL) of children and the Social Competencies Check List (SCCL) in both the 3- and 4-year-old groups did not show an impact of Head Start, i.e., the scores on the SSPAL and the SCCL scales did not statistically differ between children in the Head Start group compared to children in the non-Head Start group.

Impact on Problem Behavior

As shown in Exhibit 6.1-A for children from all language groups in the 3-year-old group, the score on the Total Behavior Problems scale at the end of the program year was 0.5 points **lower** for children in the Head Start group compared to children in the non-Head Start group. The score on the Hyperactive Behavior subscale of the Total Behavior Problems scale was also significantly lower by 0.3 points for Head Start children compared to the non-Head Start children.

No overall statistically significant impacts were found for these outcomes for children in the 4-year-old group (Exhibit 6.2-A). However, among children in this age group from English-language family backgrounds (Exhibit 6.2-B), children in the Head Start group scored significantly lower on the Aggressive Behavior subscale at the end of the program year than did similar children in the non-Head Start group (a difference of 0.22 points).

Although statistically significant, the magnitude of the Head Start impact on children's problem behaviors was small (see effect sizes in Exhibits 6.3-A and 6.3-B). The effect sizes of the impact on Total Problem Behavior of children in the 3-year-old group was 13 percent of a standard deviation, and for the impact on Hyperactive Behavior, the effect size was 18 percent of a standard deviation. The effect size of the impact on the Aggressive Behavior of 4-year-olds from English-speaking families amounted to 15 percent of a standard deviation.

Moderator/Subgroup Differences

The analysis of impacts by subgroups of children and families (detailed in Appendix 6.2 and summarized in Exhibits 6.3-A and 6.3-B for those found to be statistically significant) show some variations in impact for particular types of Head Start participants. The most important, and consistent, of these findings are discussed below, related to impacts on particular subgroups. The subgroup-specific impact findings for social-emotional outcomes, although less widespread than for the cognitive outcomes, show that children from different language and racial/ethnic groups are benefiting from Head Start.

Other statistically significant findings in Exhibits 6.3-A and 6.3-B are not discussed because it is possible they are due to chance alone and do not represent true impacts of the intervention (see discussion of subgroup impact analysis in Chapter 4).[1]

Impacts on Particular Subgroups

Child language. Among children in the 3-year-old group, significant negative impacts of Head Start on Total Problem Behavior were found for the subgroup of children from English-speaking families (i.e., Head Start reduced the incidence of reported problem behaviors), while significant negative impacts on Hyperactive Behavior were found separately for children from English-speaking families and for children from Spanish-speaking families (i.e., reductions in the reported incidence of aggressive behavior). Among children in the 4-year-old group, the significant negative impact of Head Start on Aggressive Behavior was found only in the subgroup of children from English-speaking families. No other differences were found to be statistically significant for either language group.

Child race/ethnicity. Among children in the 3-year-old group, there is a negative impact of Head Start on the reported Social Competencies Checklist, i.e., the parents of African American children in the Head Start group reported less social development for their children than did parents in the non-Head Start group. For Hispanic and White children in this same age group, positive impacts were found in the area of Hyperactive Behavior (i.e., a reported lower incidence of such behavior). Among African American children in the 4-year-old group,

[1] While each of the remaining subgroup findings taken one at a time is structured to limit the probability of a "false positive" to 1 in 20, as a group it is almost inevitable that some of these results will reach that level by chance alone. Only when a substantial share of all the tests of impact conducted for a given subgroup—or of a difference in impact between two subgroups—is statistically significant across all four of the outcome domains considered (not simply the outcomes reported in this chapter) can we be sure that at least some of those findings represent real impacts.

significant impacts of Head Start were found on reported Total Problem Behavior and on the reported Aggressive Behavior subscale (i.e., indicating reductions in reported negative behaviors).[2]

[2] Statistically significant findings in Exhibits 6.3-A and 6.3-B for differences in impact are not discussed because it is possible they are due to chance alone and do not represent true impacts of the intervention (see discussion of subgroup impact analysis in Chapter 4).

Exhibit 6.1-A: Initial One-Year Estimates of the Impact of Access to Head Start on Social-Emotional Outcomes: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group (Weighted Data)

Outcome Measure	Intent-To-Treat Impact Estimates				
	Head Start Mean	**Non-Head Start Mean**	**Mean Difference[2]**	**Regression-Adjusted Impact Estimates (Demographic Covariates Only)**	**Regression-Adjusted Impact Estimates (With Fall Measure)**
(Sample N=2,071) :					
Social Skills Scale	12.4	12.4	0.1	0.06	-0.0
Total Problem Behavior Scale	5.8	6.3	-0.5*	-0.49*	-0.48** (-0.13)
Aggressive Behavior Scale	3.0	3.0	-0.1	-0.07	-0.12
Hyperactive Behavior Scale	1.7	2.0	-0.3**	-0.34***	-0.29** (-0.18)
Withdrawn Behavior Scale	0.6	0.6	-0.0	-0.03	-0.03
Social Competencies Checklist	11.0	11.0	-0.0	-0.04	-0.06

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.
[1] Fall measure used in regression failed statistical test.
[2] Differences are rounded to the nearest 0.1.
Note: Numbers in parentheses in shaded boxes are estimated effect sizes.

Exhibit 6.1-B: Initial One-Year Estimates of the Impact of Access to Head Start on Social-Emotional Outcomes: 3-Year-Old Group, Fall English-Spring English Group (Weighted Data)

Outcome Measure	Intent-To-Treat Impact Estimates				
	Head Start Mean	**Non-Head Start Mean**	**Mean Difference[2]**	**Regression-Adjusted Impact Estimates (Demographic Covariates Only)**	**Regression-Adjusted Impact Estimates (With Fall Measure)**
(Sample N=1,629):					
Social Skills Scale[1]	12.4	12.3	0.1	0.14	0.06
Total Problem Behavior Scale	5.5	6.0	-0.5*	-0.53*	-0.45* (-0.12)
Aggressive Behavior Scale	2.9	3.0	-0.1	-0.11	-0.15
Hyperactive Behavior Scale[1]	1.6	1.8	-0.3*	-0.29** (-0.18)	-0.20
Withdrawn Behavior Scale	0.5	0.6	-0.1	-0.10	-0.09
Social Competencies Checklist	10.9	10.9	-0.1	-0.05	-0.08

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.
[1] Fall measure used in regression failed statistical test.
[2] Differences are rounded to the nearest 0.1.
Note: Numbers in parentheses in shaded boxes are estimated effect sizes.

Exhibit 6.1-C: Initial One-Year Estimates of the Impact of Access to Head Start on Social-Emotional Outcomes: 3-Year-Old Group, Fall Spanish-Spring English Group (Weighted Data)

Outcome Measure	Intent-To-Treat Impact Estimates				
	Head Start Mean	**Non-Head Start Mean**	**Mean Difference[2]**	**Regression-Adjusted Impact Estimates (Demographic Covariates Only)**	**Regression-Adjusted Impact Estimates (With Fall Measure)**
(Sample N=442):					
Social Skills Scale[1]	12.3	12.5	-0.2	-0.16	-0.19
Total Problem Behavior Scale[1]	6.8	7.4	-0.5	-0.56	-0.68
Aggressive Behavior Scale[1]	3.3	3.3	0.0	-0.06	-0.06
Hyperactive Behavior Scale[1]	2.2	2.8	-0.6**	-0.62** (-0.39)	-0.67**
Withdrawn Behavior Scale[1]	0.8	0.6	0.2	0.24	0.19
Social Competencies Checklist[1]	11.3	11.2	0.0	0.14	0.09

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.
[1] Fall measure used in regression failed statistical test.
[2] Differences are rounded to the nearest 0.1.
Note: Numbers in parentheses in shaded boxes are estimated effect sizes.

Exhibit 6.2-A: Initial One-Year Estimates of the Impact of Access to Head Start on Social-Emotional Outcomes: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group (Weighted Data)

Outcome Measure	Intent-To-Treat Impact Estimates				
	Head Start Mean	**Non-Head Start Mean**	**Mean Difference[2]**	**Regression-Adjusted Impact Estimates (Demographic Covariates Only)**	**Regression-Adjusted Impact Estimates (With Fall Measure)**
(Sample N= 1,638) :					
Social Skills Scale	12.5	12.5	-0.0	-0.00	-0.06
Total Problem Behavior Scale[1]	5.6	5.8	-0.3	-0.25	-0.01
Aggressive Behavior Scale[1]	2.7	2.9	-0.2	-0.14	-0.04
Hyperactive Behavior Scale[1]	1.7	1.8	-0.1	-0.09	-0.00
Withdrawn Behavior Scale	0.7	0.7	-0.0	-0.05	-0.03
Social Competencies Checklist[1]	11.0	11.1	-0.0	-0.03	-0.02

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.
[1] Fall measure used in regression failed statistical test.
[2] Differences are rounded to the nearest 0.1.

Exhibit 6.2-B: Initial One-Year Estimates of the Impact of Access to Head Start on Social-Emotional Outcomes: 4-Year-Old Group, Fall English-Spring English Group (Weighted Data)

Outcome Measure	Intent-To-Treat Impact Estimates				
	Head Start Mean	**Non-Head Start Mean**	**Mean Difference[2]**	**Regression-Adjusted Impact Estimates (Demographic Covariates Only)**	**Regression-Adjusted Impact Estimates (With Fall Measure)**
(Sample N=1,130):					
Social Skills Scale[1]	12.6	12.5	0.1	0.12	0.05
Total Problem Behavior Scale[1]	5.1	5.5	-0.4	-0.39	-0.04
Aggressive Behavior Scale[1]	2.5	2.8	-0.2	-0.22* (-0.14)	-0.05
Hyperactive Behavior Scale[1]	1.5	1.6	-0.1	-0.13	0.00
Withdrawn Behavior Scale	0.6	0.7	-0.1	-0.08	-0.07
Social Competencies Checklist	11.0	11.0	-0.0	-0.02	-0.02

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.
[1] Fall measure used in regression failed statistical test.
[2] Differences are rounded to the nearest 0.1.
Note: Numbers in parentheses in shaded boxes are estimated effect sizes.

Exhibit 6.2-C: Initial One-Year Estimates of the Impact of Access to Head Start on Social-Emotional Outcomes: 4-Year-Old Group, Fall Spanish-Spring English Group (Weighted Data)

Outcome Measure	Intent-To-Treat Impact Estimates				
	Head Start Mean	**Non-Head Start Mean**	**Mean Difference[2]**	**Regression-Adjusted Impact Estimates (Demographic Covariates Only)**	**Regression-Adjusted Impact Estimates (With Fall Measure)**
(Sample N= 508):					
Social Skills Scale	12.2	12.5	-0.4	-0.32	-0.33
Total Problem Behavior Scale	6.8	6.7	0.1	0.20	0.25
Aggressive Behavior Scale[1]	3.2	3.1	0.1	0.13	0.12
Hyperactive Behavior Scale	2.3	2.3	0.0	0.06	0.06
Withdrawn Behavior Scale[1]	0.7	0.7	0.0	0.02	0.05
Social Competencies Checklist[1]	11.0	11.1	-0.1	-0.01	0.01

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.
[1] Fall measure used in regression failed statistical test.
[2] Differences are rounded to the nearest 0.1.

Exhibit 6.3-A: Initial Estimates of the Impact of Head Start on Social Emotional Outcomes, Statistically Significant Results Only, 3-Year-Old Group, Combined English-English and Spanish-English Group (Weighted Data)

Outcome Measure	**Estimated Impact of Access to Head Start**	**Effect Size**
Overall Impact		
Total Problem Behavior Scale	-0.48**	-0.13
Hyperactive Behavior Scale	-0.29**	- 0.18
Difference in Impact [1]		
Social Competencies Checklist: Race (White Impact Exceeds African American)	0.49**	0.39
Social Competencies Checklist: Race (Hispanic Impact Exceeds African American)	0.37*	0.30
Impact on Subgroup [2]		
Total Problem Behavior Scale: No Special Needs	-0.52*	-0.14
Total Problem Behavior Scale: White	-0.86**	-0.23
Total Problem Behavior Scale: Parents Not Separated or Divorced	-0.50**	-0.13
Total Problem Behavior Scale: Parent Not Married	-0.47*	-0.13
Total Problem Behavior Scale: English-English Language Group	-0.46*	-0.12
Aggressive Behavior: White	-0.30*	-0.17
Hyperactive Behavior Scale: No Special Needs	-0.30*	-0.19
Hyperactive Behavior Scale: White	-0.34*	-0.22
Hyperactive Behavior Scale: Hispanic	-0.40*	-0.25
Hyperactive Behavior Scale: Male	-0.31*	-0.19
Hyperactive Behavior Scale: Parent Not Separated or Divorced	-0.33***	-0.21
Hyperactive Behavior Scale: Parent Married	-0.39*	-0.25
Hyperactive Behavior Scale: Parent Not Married	-0.25*	-0.16
Hyperactive Behavior Scale: English-English Language Group	-0.20*	-0.13
Hyperactive Behavior Scale: Spanish-English Language Group	-0.68**	-0.43
Social Competencies Checklist: African American	-0.34**	-0.27

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

[1] A total of 60 differences in impacts between subgroups were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 6.2. Findings for depression indicate the change in Head Start's estimated impact that accompanies a 1-point increase in mother's baseline depression score. Findings for baseline factors other than depression indicate the amount by which Head Start's estimated impact for the first subset of participants listed in the row label exceeds that for the second subset listed.

[2] A total of 78 subgroup impacts were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 6.2.

Exhibit 6.3-B: Initial Estimates of the Impact of Head Start on Social Emotional: Statistically Significant Results Only, 4-Year-Old Group, Combined English-English Spanish-English Group (Weighted Data)

Outcome Measure	Estimated Impact of Access to Head Start	Effect Size
Overall Impact		
No statistically significant impacts	N/A	N/A
Difference in Impact [1]		
Social Competencies Checklist: Depression	-0.00*	- 0.00
Aggressive Behavior: (African American Impact Exceeds Hispanic)	0.81**	0.51
Impact on Subgroup [2]		
Total Problem Behavior Scale: African American	-0.92**	-0.27
Aggressive Behavior Scale: African American	-0.61**	-0.38
Aggressive Behavior Scale: Female	-0.30*	-0.19
Aggressive Behavior Scale: English-English Language Group	-0.24*	-0.15

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

[1] A total of 60 differences in impacts between subgroups were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 6.2. Findings for depression indicate the change in Head Start's estimated impact that accompanies a 1-point increase in mother's baseline depression score. Findings for baseline factors other than depression indicate the amount by which Head Start's estimated impact for the first subset of participants listed in the row label exceeds that for the second subset listed.

[2] A total of 78 subgroup impacts were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 6.2.

Chapter 7: Impact of Head Start on Children's Health Status and Access to Health Services

Highlights

By the end of the program year, Head Start had positive, albeit modest, average impacts on some indicators of children's health:

- For children in both the 3- and 4-year-old group, a relatively large and statistically significant impact was found on the receipt of dental care, i.e., Head Start children were more likely to have received dental care than non-Head Start children.
- For children in the 3-year-old group, a statistically significant impact was found on parents' reported ratings of their children's health status, i.e., more parents of children in the Head Start group reported that their child's health was either excellent or very good.
- There were several statistically significant impacts of Head Start for children in both age groups whose native language is not English, including, for children in the 3-year-old group, positive impacts on parental reports of their child's health status and on the receipt of dental care. For children in the 4-year-old group, there was a significant impact on whether the child had health insurance and on the receipt of dental care.
- Related to the findings on home language, significant impacts are found for Hispanic children in both age groups on receipt of dental care, and for children in the 3-year-old group on parental reports of their child's health status.
- Among children in both the 3- and 4-year-old groups, positive impacts were found on parental reports of child's health status for children with special needs and on the receipt of dental care for children with special needs in the 3-year-old group.
- The impact of Head Start on children's receipt of dental care was found to **increase** with increasing levels of reported caregiver depression at baseline for children in the 3-year-old group. Among children in the 3-year-old group, the positive impact of Head Start on parent's report of their child's health status also increased with higher levels of reported initial caregiver depression.
- Among children whose parents were married, those in the Head Start group were rated higher by their parents on health status than those in the non-Head Start group, for the 3-year-old group.

It is important to note that the analysis of Head Start's impact on children's health is based solely on reports from parents. No direct measurement of children's actual health status, or their receipt of health care services, was undertaken for this study.

Organization and Presentation of Findings

This chapter focuses on the impact of Head Start on a few selected measures of children's health as reported by parents in spring 2003. As described in Chapter 4, these measures included parent report of whether the child had health insurance or dental care, the child's health status, and whether the child needed ongoing medical care in general or for an injury.

As in previous chapters, the discussion is based on an examination of statistically significant "intent-to-treat" impact estimates using the complete sample of children who were randomly assigned in 2002, focusing first on overall average impacts for all newly entering children in both the 3- and 4-year-old groups. The discussion then moves to an examination of the extent to which impacts occurred for key subgroups of Head Start children and how different in size impacts may be for various subgroups. Appendix 7.1 presents estimated impacts of Head Start on program participants.

The statistical results discussed in this chapter are presented in a series of tables, some of which are provided in Appendix 7.2. Exhibits 7.1 (for children in the 3-year-old group) and 7.2 (for the 4-year-old group) present the overall average impact estimates for the combined sample. Exhibits 7.3-A and 7.3-B (for the 3- and 4-year-old groups, respectively) summarize all of the statistically significant average impacts (both for the overall group and for a set of 10 subgroups discussed in Chapter 4) and provide both the estimated impact and its associated effect size. Finally, Exhibits 7.4 through 7.13, provided in Appendix 7.2, show the complete set of results of the moderator/subgroup analyses, with a separate table for each individual measure of health outcomes (again, only for the full combined sample).

Estimated Impact of Access to Head Start

This first section discusses the estimated impact of Head Start on health outcomes using the full sample of children randomly assigned to either Head Start or to the non-Head Start group, referred to as "intent-to-treat" impact estimates. These measures show the average effect of access to the program.

As shown in Exhibit 7.1, for children in the 3-year-old group, a small statistically significant impact was found for parent reports of the child's health status being excellent or very good (as shown in Exhibit 7.3-A, effect size=0.12), and a modest significant impact on the receipt of dental care (a 17 percentage point difference, effect size=0.34). As shown in Exhibit 7.2, a modest statistically significant impact was also found on the receipt of dental care for children in the 4-year-old group (a 16 percentage point difference, effect size=0.32).

The consistent, and relatively large, impact on children's receipt of dental care is particularly important in light of numerous studies that have documented substantial disparities in the level of dental services received by low-income and minority children, who are most at risk of having untreated cavities compared with other children. For example, a Government Accounting Office (GAO) study published in 2000 reported that among children ages 2 through 5 who had family incomes below $10,000, nearly one in three had at least one decayed tooth that had not been treated.[1] In contrast, only 1 in 10 preschool children whose family incomes were $35,000 or higher had untreated cavities.

This disparity is recognized in the Healthy People 2010 objectives, one of which is to "*Increase the proportion of low-income children and adolescents who received any preventive dental service during the past year*"[2] from 20 percent in 1996 (baseline) to 57 percent in 2010. The proportion of Head Start children who had received dental care exceeded the target in the Healthy People 2010 dental care objective.

Moderator/Subgroup Differences

The analysis of impacts by subgroups of children and families (detailed in Appendix 7.2 and summarized in Exhibits 7.3-A and 7.3-B for those found to be statistically significant) show some variations in impact for particular types of Head Start participants. The most notable findings are discussed below as in previous chapters beginning with possible differences in impact between or among subgroups and then examining impacts on particular subgroups.

Differences in Impact

The impact of Head Start on children's receipt of dental care was found to **increase** with increasing levels of reported caregiver depressive symptoms for children in the 3- and 4 -year-old groups. In addition, for children in the 3-year-old group, the positive impact of Head Start on parent's report of their child's health status (as good or excellent) also increased with higher levels of reported caregiver depressive symptoms. Moreover, for the 3-year-old group, Head Start had a greater impact on non-English speaking parents' report of their child's health status as good or excellent.

---

[1] US GAO. (April 2000). ***Oral Health: Dental Disease is a Chronic Problem Among Low-Income Populations.*** Washington, DC: GAO/HEHS-00-72.

[2] US Department of Health and Human Services. (2000). ***Healthy People 2010: 21 – Oral Health***. Retrieved from: www.healthypeople.gov/document/HTML/Volume2/21Oral.htm

Other statistically significant findings in Exhibits 7.3-A and 7.3-B are not discussed because it is possible they are due to chance alone and do not represent true impacts of the intervention (see discussion of subgroup impact analysis in Chapter 4).[3]

Impacts on Particular Subgroups

Home language. There were several statistically significant impacts of Head Start for children **whose home language was not English**. Among children in both the 3- and 4-year-old groups, positive impacts were found on parental reports of their child's receipt of dental care. For children in the 4-year-old group, there was also a significant impact on whether the child had health insurance, with non-English-speaking children in the Head Start group being more likely to have health insurance than similar children in the non-Head Start group.

Parent's report of their child's health status, however, provided mixed results by age group. For children in the 3-year-old group, Head Start had a positive impact on non-English-speaking families' report on health status (i.e., parents of children in the Head Start group were more likely to report their child's health as good or excellent, compared to parents in the non-Head Start group), but the opposite was true for parents of children in the 4-year-old group. Non-English-speaking parents of children in the Head Start group were **less** likely to report their child's health as being good or excellent. For children in both age groups, Head Start had a positive effect on English-speaking and non-English-speaking parents' report of their child's receipt of dental care.

Race/ethnicity. Related to the findings on home language, significant positive impacts were found for Hispanic children on several of the health measures. Significant impacts were found for children in **both** age groups on receipt of dental care and, for children in the 3-year-old group, on parental reports of their child's health status. In addition, there was an impact on White children's receipt of dental care for both age groups.

Special needs. There were also statistically significant impacts of Head Start for children in the 3-year-old group with special needs, i.e., positive impacts were found on parental reports of their child's health status and on the receipt of dental care.

---

[3] While each of the remaining subgroup findings taken one at a time is structured to limit the probability of a "false positive" to 1 in 20, as a group it is almost inevitable that some of these results will reach that level by chance alone. Only when a substantial share of all the tests of impact conducted for a given subgroup—or of a difference in impact between two subgroups—is statistically significant across all four of the outcome domains considered (not simply the outcomes reported in this chapter) can we be sure that at least some of those findings represent real impacts.

The subgroup-specific impact findings indicate widespread effects with children from all but one of the examined subgroups found to be benefiting from Head Start.

Exhibit 7.1: Initial One-Year Estimates of the Impact of Access to Head Start on Health Outcomes: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group (Weighted Data)

Outcome Measure	Intent-To-Treat Impact Estimates				
	Head Start Mean (%)	**Non-Head Start Mean (%)**	**Mean Difference[2] (%)**	**Regression-Adjusted Impact Estimates (Demographic Covariates Only) (%)**	**Regression-Adjusted Impact Estimates (With Fall Measure) (%)**
(Sample N= 2,071):					
Child Has Health Insurance	92.1	91.6	0.0	0.00	-0.00
Child Health Status Is Excellent or Very Good	80.6	75.8	5.0	6.00*	5.00* (0.12)
Child Needs Ongoing Care[1]	13.2	12.9	0.3	-0.00	0.20
Child Had Care for Injury in Last Month	9.0	8.2	0.7	0.00	0.00
Child Had Dental Care[1]	68.9	51.8	17.0***	17.00*** (0.34)	13.00***

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.
[1] Fall measure used in regression failed statistical test.
[2] Differences are rounded to the nearest 0.1.
Note: Numbers in parentheses in shaded boxes are estimated effect sizes.

Exhibit 7.2: Initial One-Year Estimates of the Impact of Access to Head Start on Health Outcomes: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group (Weighted Data)

Outcome Measure	Intent-To-Treat Impact Estimates				
	Head Start Mean (%)	**Non-Head Start Mean (%)**	**Mean Difference[2] (%)**	**Regression-Adjusted Impact Estimates (Demographic Covariates Only) (%)**	**Regression-Adjusted Impact Estimates (With Fall Measure) (%)**
(Sample N= 1,638):					
Child has Health Insurance	88.9	88.0	0.1	0.01	0.02
Child Health Status Is Excellent or Very Good	79.1	81.1	-2.0	-0.03	-0.03
Child Needs Ongoing Care[1]	11.2	11.2	0.0	0.00	0.02
Child Had Care for Injury in Last Month	11.6	12.0	-0.4	-0.01	-0.01
Child Had Dental Care[1]	73.2	56.9	16.3**	0.16** (0.32)	0.13**

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.
[1] Fall measure used in regression failed statistical test.
[2] Differences are rounded to the nearest 0.1.
Note: Numbers in parentheses in shaded boxes are estimated effect sizes.

Exhibit 7.3-A: Initial Estimates of the Impact of Head Start on Health Outcomes: Statistically Significant Results Only, 3-Year-Old Group, Combined English-English and Spanish-English Group (Weighted Data)

Outcome Measure	Estimated Impact of Access to Head Start	Effect Size
Overall Impact		
Child Health Status Excellent or Very Good	0.05*	0.12
Child Had Dental Care	0.17***	0.34
Difference in Impact [1]		
Child Health Status: Home Language (Non-English Impact Exceeds English)	0.12*	0.28
Child Health Status: Depression	0.05*	0.12
Child Had Care for Injury: Race (White Impact Exceeds African American)	0.08*	0.30
Child Had Care for Injury: Race (White Impact Exceeds Hispanic)	0.13***	0.48
Child Had Dental Care: Depression	0.16***	0.32
Impact on Subgroup [2]		
Child Health Status: Special Needs	0.19*	0.44
Child Health Status: Parent Married	0.08*	0.19
Child Health Status: Hispanic	0.12**	0.28
Child Health Status: Home Language Not English	0.14**	0.33
Child Had Care for Injury: White	0.07***	0.26
Child Had Care for Injury: Hispanic	-0.06*	- 0.22
Child Had Dental Care: Special Needs	0.24*	0.48
Child Had Dental Care: No Special Needs	0.16***	0.32
Child Had Dental Care: Parent Married	0.18***	0.36
Child Had Dental Care: Parent Not Married	0.16***	0.32
Child Had Dental Care: White	0.17***	0.34
Child Had Dental Care: Hispanic	0.22***	0.44
Child Had Dental Care: Home Language Not English	0.22***	0.44
Child Had Dental Care: Home Language English	0.15***	0.30

* = p≦0.05, ** = p≦0.01, *** = p≦0.001.

[1] A total of 35 differences in impacts between subgroups were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 7.2. Findings for depression indicate the change in Head Start's estimated impact that accompanies a 1-point increase in mother's baseline depression score. Findings for baseline factors other than depression indicate the amount by which Head Start's estimated impact for the first subset of participants listed in the row label exceeds that for the second subset listed.

[2] A total of 50 subgroup impacts were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 7.2.

Exhibit 7.3-B: Initial Estimates of the Impact of Head Start on Health Outcomes: Statistically Significant Results Only, 4-Year-Old Group, Combined English-English Spanish-English Group (Weighted Data)

Outcome Measure	Estimated Impact of Access to Head Start	Effect Size
Overall Impact		
Child Had Dental Care	0.16***	0.32
Difference in Impact [1]		
Child Had Health Insurance: Race (Hispanic Impact Exceeds African American)	0.08*	0.24
Child Health Status: Special Needs (No Special Needs Impact Exceeds Special Needs)	0.22*	0.56
Child Had Dental Care: Depression	0.16***	0.32
Impact on Subgroup [2]		
Child Had Health Insurance: Home Language Not English	0.06*	0.18
Child Health Status: Special Needs	-0.23*	-0.59
Child Health Status: Parent Married	-0.08**	-0.21
Child Health Status: Home Language Not English	-0.08*	-0.21
Child Had Dental Care: No Special Needs	0.16***	0.32
Child Had Dental Care: Parent Married	0.18***	0.36
Child Had Dental Care: Parent Not Married	0.14**	0.28
Child Had Dental Care: White	0.24***	0.48
Child Had Dental Care: Hispanic	0.12*	0.24
Child Had Dental Care: Home Language Not English	0.17**	0.34
Child Had Dental Care: Home Language English	0.16***	0.32

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

[1] A total of 35 differences in impacts between subgroups were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 7.2. Findings for depression indicate the change in Head Start's estimated impact that accompanies a 1-point increase in mother's baseline depression score. Findings for baseline factors other than depression indicate the amount by which Head Start's estimated impact for the first subset of participants listed in the row label exceeds that for the second subset listed.

[2] A total of 50 subgroup impacts were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 7.2.

Chapter 8: Impact of Head Start on Parenting Practices

Highlights

By the end of the program year, Head Start had positive, albeit modest, average impacts on parenting practices:

- For both age cohorts, Head Start had a consistently positive overall average impact on the amount of time parents reported reading to their child, with parents of Head Start children significantly more likely to read to their child than parents of non-Head Start children. Statistically significant average impacts were also found for 3-year-olds on the extent to which their parents exposed them to a variety of cultural enrichment activities such as taking them to a museum or a zoo.
- For parents of 3-year-olds, there is a small, but statistically significant, reduction in the use of physical discipline, but no impact was found for parents of children in the 4-year-old group. Parents of children in the Head Start group were significantly less likely than non-Head Start parents to report using spanking when their child misbehaved and reported using it less frequently.
- No statistically significant impacts were found on parents' child safety practices at home, for parents of children in both the 3- and 4-year-old group.
- Significant impacts were also found for specific subgroups of parents and/or children.
    - For the 3-year-old group, mothers who had first given birth before age 19 had significant impacts in the area of physical discipline, while significant impacts for mothers who had first given birth after age 19 were found in the area of educational activities.
    - Among parents of children in the 3-year-old group, the impact of Head Start on the **use** of physical discipline (i.e., spanking) **decreased** with increasing levels of depressive symptoms, but Head Start's impact on the **frequency** of physical discipline **increased** with increasing levels of depressive symptoms.
    - Parents of 3-year-olds whose primary language was English were especially likely to benefit from Head Start, with significant impacts on the likelihood of reading to their child, on a reduced use of spanking and the frequency of its use, and on the use of child safety practices.
    - For parents of boys in the 3-year-old group, there is a significant reduction in parent's use of spanking as a disciplinary strategy.

Introduction

This chapter shifts focus from impacts on children to the potential positive benefits of Head Start for low-income parents. One of the hallmarks of Head Start is its recognition that parents are their child's first and primary teacher and that the involvement of parents is crucial for fostering children's school readiness.[1] From the beginning, Head Start programs have reached out to families in a variety of ways, by encouraging parent involvement in their child's classroom, providing parent education to help strengthen parents' childrearing knowledge and skills, and providing referrals to address family needs so that parents can be more effective in their role as caregiver.

A strong, nurturing parent-child relationship is essential for healthy cognitive and social-emotional growth during early childhood.[2] Parent-child interactions that involve talking, reading, teaching, and exposure to new experiences are crucial for promoting language development and early literacy. Parents can also support their young child's cognitive development by providing a stimulating learning environment at home and in the community. Parental discipline that emphasizes establishing firm but fair expectations for child behavior promotes the development of social understanding and skills necessary for positive relationships with peers and adults. Parental nurturance provides young children with the emotional support needed for developing trusting relationships with adults, learning to regulate their emotional responses, and playing cooperatively with peers. Finally, parents' preventive efforts to safeguard the child's environment are crucial for children's physical health and overall well-being. Head Start's efforts to support parents on these dimensions of childrearing can go a long way in ensuring that the Head Start services that children receive are complemented and augmented by what their experiences at home.

Organization and Presentation of Findings

The measures used in this report to assess the impact of Head Start on childrearing practices focus on three key parenting constructs—educational activities, discipline strategies, and child safety practices. Selection of these measures was guided by several factors, including relevance for program goals, appropriateness for Head Start families, prior use in national studies

---

[1] Zigler, E., & S. Muenchow. (1992). ***Head Start: The Inside Story of America's Most Successful Educational Experiment.*** New York: Basic Books.

[2] National Research Council. (2000). ***From Neurons To Neighborhoods: The Science of Early Childhood Development***. Washington, DC: National Academy Press.

and evaluations, and adequate psychometric properties. These skill-based dimensions of childrearing emphasizing cognitive stimulation, child discipline, and child safety are common elements of parent education offered through Head Start and thus are likely to be affected by parents' access to the program. Prior research with similar populations has shown significant associations between these domains and children's cognitive and social-emotional development[3]. Moreover, these domains are likely to be important mediators of the impact of Head Start on children's health and development. As noted in Chapter 4, the specific measures of childrearing practices[4] examined in this report cover three constructs—educational activities, disciplinary practices, and safety practices.

The target sample for these analyses included all caregivers who identified themselves in spring 2003 as the person primarily responsible for the study child's daily care and overall well-being. For the vast majority of study children (92%), the primary caregiver was the child's biological or adoptive mother.[5] For simplicity of discussion, throughout the chapter we use the term "parent" when referring to the primary caregiver.

As in previous chapters, this discussion of estimated impacts examines statistically significant "intent-to-treat" impact estimates using the complete sample of children who were randomly assigned in 2002, focusing first on overall average impacts for all newly entering children in both the 3- and 4-year-old groups and then examining any notable differences in average impacts by the language used for child assessment. The discussion then moves to an examination of the extent to which impacts occurred for key subgroups of Head Start children,

---

[3] Administration on Children, Youth, and Families. (2003). ***Head Start FACES 2000: A Whole Child Perspective On Program Performance, Fourth Progress Report***. Washington, DC: US Department of Health and Human Services; Administration on Children, Youth, and Families. (2002). ***Making a Difference in the Lives of Infants and Toddlers and Their Families: The Impacts of Early Head Start.*** Washington, DC: US Department of Health and Human Services.

[4] It is important to note that all of the childrearing measures used in this analysis are based on parent reports of their own, or other family members', behavior and are therefore susceptible to response biases inherent in self-reported data. In the absence of observations of parent-child interactions, or other reporter data (e.g., interviewer assessments of the home environment), it is difficult to determine the degree of response bias and whether it represents an over- or under-estimate of parents' actual childrearing practices. Therefore, caution should be used when interpreting obtained group means and proportions as reflecting actual levels of childrearing practices. However, the random assignment design of the Head Start Impact Study ensures that the degree and direction of bias should, on average, be equivalent for all families regardless of their assignment to the treatment or control group, i.e., the Head Start—non-Head Start difference is unbiased.

[5] Four percent of the caregivers were biological or adoptive fathers, 3 percent were grandparents, and about 1 percent were either other relatives or individuals unrelated to the study child. In 95 percent of study families, the individual identified as the child's primary caregiver in spring 2003 was also identified as the child's primary caregiver at baseline in fall 2002. In the remaining 5 percent of cases, the primary caregiver assumed this responsibility at some point between the fall and spring assessments.

and how different in size impacts may be for various subgroups. Appendix 8.1 presents estimated impacts on program participants.

The statistical results discussed in this chapter are presented in a series of tables, some of which are provided in Appendix 8.2. Exhibits 8.1 (for children in the 3-year-old group) and 8.2 (for the 4-year-old group), present the overall average impact estimates for the combined sample. Exhibits 8.3-A and 8.3-B (for the 3- and 4-year-old groups, respectively) summarize all of the statistically significant average impacts (both for the overall group and for a set of 12 subgroups discussed in Chapter 4) along with their associated effect sizes. Finally, Exhibits 8.4 through 8.23, provided in Appendix 8.2, show the results of the moderator/subgroup analyses, with a separate table for each individual measure of parenting outcomes.

Estimated Impact of Access to Head Start

This first section discusses the estimated impact of Head Start on parenting practices outcomes using the sample of children randomly assigned to either Head Start or to the non-Head Start group, referred to as "intent-to-treat" impact estimates. These measures show the average impact of access to the program.

Educational Activities

As shown in Exhibits 8.1 and 8.2, for children in both age groups, small but statistically significant **positive** impacts were found for parents' reading to their child (effect size = 0.18 and 0.13 for children in the 3- and 4-year-old groups, respectively). These results are consistent with program impacts of similar magnitude from the National Evaluation of Early Head Start regarding daily reading among parents of 3-year-olds. They are also encouraging in light of accumulating evidence that the amount of shared reading at home plays a critical role in low-income children's language development and emergent literacy.[6]

Statistically significant impacts were also found for children in the 3-year-old group on the extent to which their parents exposed them to a variety of cultural enrichment activities, with Head Start parents providing significantly more enrichment activities for their child than parents of non-Head Start children by the end of the program year (an effect size of 11 percent). Although the impact of Head Start on this aspect of childrearing is small (and not detected for parents of children in the 4-year-old group), it does indicate that, at least for the younger children, Head

[6] Bus, A.G., M.H. van IJzendoorn, & A. Pellegrini. (1995). "Joint Book Reading Makes for Success in Learning to Read: A Meta-Analysis on Intergenerational Transmission of Literacy." ***Review of Educational Research, 65***, 1-21.

Start parents are making greater efforts to broaden their child's world to include learning experiences like trips to the zoo, local museums, and cultural events.

Disciplinary Practices

Small, but statistically significant, impacts were found on the physical disciplinary practices used by parents of children in the 3-year-old group; however, no statistically significant impacts were found on physical disciplinary practices for children in the 4-year-old group. By the end of the first program year, parents of children in the Head Start group were significantly less likely than non-Head Start parents to report the use of spanking in the last week (an effect size of -14 percent) and the frequency with which spanking was used during the past week (an effect size of -10 percent) when their child misbehaved. These results are consistent with findings from the National Evaluation of Early Head Start, which found significantly lower reported use of physical punishment among program parents. They are also promising in light of evidence from experimental studies that interventions to reduce parents' reliance on physical discipline to gain child compliance can lead to improvements in Head Start children's social behavior at home and in preschool.[7]

Safety Practices

No statistically significant impacts were found on parents' child safety practices at home. Most Head Start and non-Head Start parents of children in both the 3- and 4-year-old groups reported "almost always" or "always" storing medicines and cleaning supplies out of children's reach, supervising the child during bath time, using a child car seat, and following other safety practices. These results suggest that low-income parents may already have a strong awareness of what is needed to protect their child from harm, coming from other sources such as their pediatrician, family and friends, or media campaigns. Of course, it must be kept in mind that these are reported data and may differ from parent's actual behavior.

Moderator/Subgroup Differences

As in the previous chapters, the impact of Head Start on childrearing was examined for key subgroups of parents, acknowledging that Head Start may be especially beneficial for certain

---

[7] Webster-Stratton, C. (1998). "Preventing Conduct Problems in Head Start Children: Strengthening Parent Competencies." ***Journal of Consulting and Clinical Psychology, 66*****,** 715-730; Webster-Stratton, C., M.J. Reid, & M. Hammond. (2001). "Preventing Conduct Problems and Promoting Social Competence: A Parent and Teacher Training Partnership in Head Start." ***Journal of Clinical Child Psychology, 30*****,** 283-302.

types of caregivers. These analyses, detailed in Appendix 8.2 and summarized in Exhibits 8.3-A and 8.3-B for those found to be statistically significant, show some variations in impact for particular types of parents. The most notable findings are discussed below in the same way as in previous chapters.

Other statistically significant findings in Exhibits 8.3-A and 8.3-B are not discussed because it is possible they are due to chance alone and do not represent true impacts of the intervention (see discussion of subgroup impact analysis in Chapter 4).[8]

Differences in Impact

A statistically significant relationship was found between the impact of Head Start on childrearing practices and reported parental depression for children in the 3-year-old group. However, the results are mixed. Among parents of children in the 3-year-old group, the impact of Head Start on the **use** of physical discipline (i.e., spanking) **decreased** with increasing levels of depressive symptoms, but the **frequency** of physical discipline **increased** with increasing levels of depressive symptoms. Taken together, these findings indicate that Head Start parents with elevated depressive symptoms are significantly less likely to use physical discipline when their child misbehaves, but if they use physical discipline, use it more often than other parents.

Parents with chronic depressive symptoms tend to be less sensitive and responsive to their children's needs, less nurturing, and more erratic and punitive in their discipline practices, all of which have serious consequences for children's development and well-being. While these parents represent a highly vulnerable group, their greater needs also make them more difficult to engage and serve in intervention programs.[9] The mixed pattern of relationships between depression and program impacts may, therefore, reflect the difficulties Head Start staff face in engaging and working with this group of caregivers.

In the 3-year-old group, Head Start had a larger effect on discipline strategies (i.e. decreasing spanking) for mothers who gave birth as teenagers than those who gave birth after age 19.

---

[8] While each of the remaining subgroup findings taken one at a time is structured to limit the probability of a "false positive" to 1 in 20, as a group it is almost inevitable that some of these results will reach that level by chance alone. Only when a substantial share of all the tests of impact conducted for a given subgroup—or of a difference in impact between two subgroups—is statistically significant across all four of the outcome domains considered (not simply the outcomes reported in this chapter) can we be sure that at least some of those findings represent real impacts.

[9] Administration on Children, Youth, and Families. (2002). ***Making A Difference in the Lives of Infants and Toddlers and Their Families: The Impacts Of Early Head Start.*** Washington, DC: US Department of Health and Human Services.

Impacts on Particular Subgroups

Age of Mother at First Birth. Head Start had positive impacts on the childrearing practices of mothers who first gave birth as a teenager ("teen mothers")[10] and for mothers who had their first baby when they were older ("not teen mothers"), but the impacts are found in different areas of parenting practices for the two groups. Mothers who had first given birth before age 19 had significant impacts in the area of physical discipline, while significant impacts for mothers who had first given birth after age 19 were found in the area of educational activities.

Among parents of children in the 3-year-old group, teen mothers of Head Start children were significantly less likely than non-Head Start teen mothers to use spanking and to use it less frequently when their child misbehaved. The size of the impacts for teen mothers were sizable (subgroup effect sizes of -34 percent for the use of spanking and -23 percent for the frequency of spanking), more than twice as large as the impacts obtained for the sample overall (overall effect sizes of -14 percent and -10 percent, respectively). These results are consistent with findings from the National Evaluation of Early Head Start, which found significant program impacts on the use of physical discipline among mothers who were 19 or younger when their child was born. These findings are also encouraging in light of general consensus that those who become mothers in adolescence are at heightened risk for punitive parenting practices, as well as child abuse and neglect.[11] Although Head Start was not designed specifically to serve the needs of teenage mothers, these findings suggest that access to the program can have beneficial effects in reducing children's risk for punitive discipline practices, although results of studies of efforts to improve the parenting skills of young low-income mothers have had mixed results. [12]

Statistically significant impacts of Head Start were also found for the educational activities provided by mothers who had first given birth after age 19. Among these non-teen mothers of children in the 3-year-old group, those in the Head Start group spent significantly more time reading to their child and taking them to a greater variety of cultural enrichment activities than mothers in the non-Head Start group. A similarly positive impact of Head Start on

[10] It is important to keep in mind that this variable refers to whether the mother was ever a teen mother and not whether she gave birth to the target child as a teenager.

[11] Maynard, R. (1996). ***Kids Having Kids: Economic Costs and Social Consequences of Teen Pregnancy.*** Washington, DC: Urban Institute Press.

[12] Kisker, E., A. Rangarajan, & K. Boller. (1998). ***Moving Into Adulthood: Were the Impacts of Mandatory Programs for Welfare-Dependent Parents Sustained after the Programs Ended?*** Princeton, NJ: Mathematica Policy Research, Inc.; Qunit, J.C., J.M. Bos, & Polit, D.F. (1997). ***New Chance: Final Report on a Comprehensive Program for Young Mothers in Poverty and Their Children.*** New York: Manpower Demonstration Research Corporation.

reading to the child at home was found for the non-teen mothers of children in the 4-year-old group. These results are also consistent with findings from the National Evaluation of Early Head Start, which found significant program impacts for older, but not younger, mothers on an array of language and literacy-promoting practices.

Finally, a small, but statistically significant, impact of Head Start was found on parents' use of "time out" but only for the non-teen mothers of children in the 4-year-old group. Mothers in the Head Start group were less likely than similar non-Head Start mothers to report placing their child in time out when they misbehaved.

Home Language. Parents of 3-year-olds whose primary language was English were especially likely to benefit from Head Start. Among native English speakers, those in the Head Start group were significantly more likely to read to their child and were less likely to use spanking, and to use it less frequently, when their child misbehaved than parents in the non-Head Start group. These impacts were modest in size, but taken together suggest that Head Start may be more effective in working with native English-speaking parents than with parents with limited English language skills.

These findings highlight an important subgroup of families that could benefit from efforts tailored to their unique language and cultural needs. Limited-English-proficient parents and children, and immigrant families with children more generally, are the fastest growing segment of the nation's low-income population, with children of immigrants currently constituting a quarter of all children under age five.[13] While their need for services are as high, or higher, than those of U.S.-born families, some may be ineligible for many Federal and state public assistance benefits, and language and cultural barriers often deter them from seeking out benefits for which they or their children are eligible[14].

Gender. The most notable finding by child gender is in the area of physical discipline, with parents of boys in the 3-year-old group significantly less likely to use spanking as a disciplinary strategy. In addition, among children in the 4-year-old group, parents are less likely to use spanking for discipline for girls compared to boys. Also, it was found that parents of girls in the 3-year-old Head Start group were more likely to read to them than parents in the non-Head Start group.

---

[13] Hernandez, D., & E. Charney. (1998). ***From Generation to Generation: The Health and Adjustment of Children in Immigrant Families***. Washington, DC: National Academy Press.

[14] Fix, M., & W. Zimmerman. (1999). ***All Under One Roof: Mixed-Status Families in an Era of Reform***. Washington, DC: The Urban Institute.

Exhibit 8.1: Initial One-Year Estimates of the Impact of Access to Head Start on Parenting Outcomes: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group (Weighted Data)

Outcome Measure	Intent-To-Treat Impact Estimates				
	Head Start Mean	**Non-Head Start Mean**	**Mean Difference[2]**	**Regression-Adjusted Impact Estimates (Demographic Covariates Only)**	**Regression-Adjusted Impact Estimates (With Fall Measure)**
(Sample N=2,071):					
Number of Times Child is Read to On Average[1]	2.9	2.8	0.2*	0.17** (0.18)	0.13*
Family Cultural Enrichment Scale	3.8	3.5	0.2**	0.19*	0.15* (0.11)
Used Time Out in Last Week? [1]	0.6	0.7	-0.0	-0.03	-0.02
Number of Times Used Time Out in Last Week	1.6	1.9	-0.3	-0.23	-0.21
Spanked Child in Last Week? [1]	0.4	0.5	-0.1*	-0.07* (-0.14)	-0.06
Number Times Spanked Child in Last Week[1]	0.8	1.0	-0.2*	-0.16* (-0.10)	-0.06
Parental Safety Practices Scale[1]	3.7	3.7	0.0	0.03	0.02
Removing Harmful Objects Subscale	3.9	3.9	0.0	0.03	0.02
Restricting Child Movement Subscale[1]	3.9	3.9	-0.0	-0.02	-0.02
Safety Devices Subscale	3.4	3.3	0.1	0.07	0.05

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.
[1] Fall measure used in regression failed statistical test.
[2] Differences are rounded to the nearest 0.1.
Note: Numbers in parentheses in shaded boxes are estimated effect sizes.

Exhibit 8.2: Initial One-Year Estimates of the Impact of Access to Head Start on Parenting Outcomes: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group (Weighted Data)

Outcome Measure	Intent-To-Treat Impact Estimates				
	Head Start Mean	**Non-Head Start Mean**	**Mean Difference[2]**	**Regression-Adjusted Impact Estimates (Demographic Covariates Only)**	**Regression-Adjusted Impact Estimates (With Fall Measure)**
Sample (N=1,638):					-
Number of Times Child is Read to On Average[1]	3.0	2.8	0.2*	0.13* (0.13)	0.11
Family Cultural Enrichment Scale	4.0	3.9	0.1	0.08	0.11
Used Time Out in Last Week?	0.6	0.7	-0.1*	-0.09*	-0.10**
Number of Times Used Time Out in Last Week	1.7	1.7	0.1	0.03	0.04
Spanked Child in Last Week?	0.4	0.4	0.0	-0.01	-0.01
Number Times Spanked Child in Last Week[1]	0.7	0.7	0.0	0.02	-0.04
Parental Safety Practices Scale	3.7	3.7	0.0	0.03	0.04
Removing Harmful Objects Subscale	3.9	3.9	0.0	0.00	-0.00
Restricting Child Movement Subscale	3.9	3.9	0.0	0.02	0.02
Safety Devices Subscale	3.4	3.4	0.1	0.06	0.10

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.
[1] Fall measure used in regression failed statistical test.
[2] Differences are rounded to the nearest 0.1.
Note: Numbers in parentheses in shaded boxes are estimated effect sizes.

Exhibit 8.3-A: Initial Estimates of the Impact of Head Start on Parenting Outcomes: Statistically Significant Results Only, 3-Year-Old Group, Combined English-English and Spanish-English Group (Weighted Data)

Outcome Measure	Estimated Impact of Access to Head Start	Effect Size
Overall Impact		
Number of Times Child is Read To	0.17**	0.18
Family Cultural Enrichment Scale	0.15*	0.11
Spanked Child in Last Week	-0.07*	-0.14
Number Time Spanked Child in Last Week	-0.16*	-0.10
Difference in Impact [1]		
Spanked Child in Last Week (Teen Mom Impact Exceeds Not Teen Mom)	0.16**	0.32
Spanked Child in Last Week: Depression	-0.07*	-0.14
Number of Times Spanked Child: Depression	0.01*	0.01
Parental Safety Practices Scale: Home Language (English Impact Exceeds Not English)	0.09*	0.27
Safety Devices Subscale (English Impact Exceeds Not English)	0.22*	0.29
Impact on Subgroup [2]		
Number of Times Child is Read To: Not Teen Mom	0.16*	0.17
Number of Times Child is Read To: Female	0.23*	0.25
Number of Times Child is Read To: White	0.27*	0.29
Number of Times Child is Read To: Parent Married	0.28**	0.30
Number of Times Child is Read To: Home Language is English	0.19**	0.20
Family Cultural Enrichment Scale: Not Teen Mom	0.23**	0.16
Family Cultural Enrichment Scale: Male	0.28*	0.20
Family Cultural Enrichment Scale: Black	0.24*	0.17
Number of Time Outs in Last Week: Female	-0.32*	-0.17
Spanked Child in Last Week: Teen Mom	-0.17***	-0.34
Spanked Child in Last Week: Male	-0.11*	-0.22
Spanked Child in Last Week: Home Language English	-0.10*	-0.20
Spanked Child in Last Week: Parent Married	-0.11*	-0.22
Number of Times Spanked Child: Teen Mom	-0.36*	-0.23
Number of Times Spanked Child: Black	-0.35*	-0.22
Number of Times Spanked Child: Home Language English	-0.25**	-0.16

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.

[1] A total of 80 differences in impacts between subgroups were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 8.2. Findings for baseline factors other than depression indicate the amount by which Head Start's estimated impact for the first subset of participants listed in the row label exceeds that for the second subset listed.

[2] A total of 110 subgroup impacts were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 8.2. Findings for depression indicate the change in Head Start's estimated impact that accompanies a 1-point increase in mother's baseline depression score.

Exhibit 8.3-B: Initial Estimates of the Impact of Head Start on Parenting Outcomes: Statistically Significant Results Only, 4-Year-Old Group, Combined English-English Spanish-English Group (Weighted Data)

Outcome Measure	Estimated Impact of Access to Head Start	Effect Size
Overall Impact		
Number of Times Child is Read To	0.13**	0.13
Difference in Impact [1]		
Spanked Child in Last Week: Gender (Female Impact Exceeds Male)	0.15*	0.31
Used Time Out in Last Week: Depression	-0.09*	-0.19
Impact on Subgroup [2]		
Number of Times Child is Read To: Not Teen Mom	0.18*	0.18
Family Cultural Enrichment Scale: Hispanic	0.22*	0.15
Used Time Out in Last Week: Not Teen Mom	-0.12*	-0.26
Used Time Out in Last Week: Male	-0.12*	-0.26
Used Time Out in Last Week: White	-0.11**	-0.23
Used Time Out in Last Week: Parent Not Married	-0.08*	-0.17
Used Time Out in Last Week: Home Language English	-0.11**	-0.23
Safety Devices Subscale: Home Language Not English	0.22*	0.29

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.

[1] A total of 80 differences in impacts between subgroups were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 8.2. Findings for depression indicate the change in Head Start's estimated impact that accompanies a 1-point increase in mother's baseline depression score. Findings for baseline factors other than depression indicate the amount by which Head Start's estimated impact for the first subset of participants listed in the row label exceeds that for the second subset listed.

[2] A total of 110 subgroup impacts were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 8.2.

Appendix 1.1: Section 649(g) of the Head Start Act, 1998 (PL 105-285)

(g) NATIONAL HEAD START IMPACT STUDY.--

(1) EXPERT PANEL.--

(A) IN GENERAL.--The Secretary shall appoint an independent panel consisting of experts in program evaluation and research, education, and early childhood programs--

(i) to review, and make recommendations on, the design and plan for the research (whether conducted as a single assessment or as a series of assessments) described in paragraph (2), within 1 year after the date of enactment of the Coats Human Services Reauthorization Act of 1998;

(ii) to maintain and advise the Secretary regarding the progress of the research; and

(iii) to comment, if the panel so desires, on the interim and final research reports submitted under paragraph (7).

(B) TRAVEL EXPENSES.--The members of the panel shall not receive compensation for the performance of services for the panel, but shall be allowed travel expenses, including per diem in lieu of subsistence, at rates authorized for employees of agencies under subchapter I of chapter 57 of title 5, United States Code, while away from their homes or regular places of business in the performance of services for the panel. Notwithstanding section 1342 of title 31, United States Code, the Secretary may accept the voluntary and uncompensated services of members of the panel.

(2) GENERAL AUTHORITY: After reviewing the recommendations of the expert panel, the Secretary shall make a grant to, or enter into a contract or cooperative agreement with an organization to conduct independent research that provides a national analysis of the impact of Head Start programs. The Secretary shall ensure that the organization shall have expertise in program evaluation, and research, education, and early childhood programs.

(3) DESIGNS AND TECHNIQUES.--The Secretary shall ensure that the research uses rigorous methodological designs and techniques, (based on the recommendations of the expert panel) including longitudinal designs, control groups, nationally recognized standardized measures, and random selection and assignment, as appropriate. The Secretary may provide that the research shall be conducted as a single comprehensive assessment or as a group of coordinated assessments designed to provide, when taken together, a national analysis of the impact of Head Start programs.

(4) PROGRAMS.--The Secretary shall ensure that the study focuses primarily on Head Start programs that operate in the 50 States, the Commonwealth of Puerto Rico or the District of Columbia and that do not specifically target special populations.

(5) ANALYSIS.--The Secretary shall ensure that the organization conducting the research--

(A)(i) determines if, overall, the Head Start programs have impacts consistent with their primary goal of increasing the social competence of children, by increasing the everyday effectiveness of the children in dealing with their present environments and future responsibilities, and increasing their school readiness;

(ii) considers whether the Head Start programs--

(I) enhance the growth and development of children in cognitive, emotional, and physical health areas;

(II) strengthen families as the primary nurturers of their children; and

(III) ensure that children attain school readiness; and

(iii) examines--

(I) the impact of the Head Start programs on increasing access of children to such services as educational, health, and nutritional services, and linking children and families to needed community services; and

(II) how receipt of services described in subclause (I) enriches the lives of children and families participating in Head Start programs;

(B) examines the impact of Head Start programs on participants on the date the participants leave Head Start programs, at the end of kindergarten, and at the end of first grade (whether in public or private school), by examining a variety of factors, including educational achievement, referrals for special education or remedial course work, and absenteeism;

(C) makes use of random selection from the population of all Head Start programs described in paragraph (4) in selecting programs for inclusion in the research; and

(D) includes comparisons of individuals who participate in Head Start programs with control groups (including control groups) composed of--

(i) individuals who participate in other early childhood programs (such as public or private preschool programs and day care); and

(ii) individuals who do not participate in any other early childhood program; and

(6) CONSIDERATION OF SOURCES OF VARIATION.--In designing the research, the Secretary shall, to the extent practicable, consider addressing possible sources of variation in impact of Head Start programs, including variations in impact related to such factors as—

(A) Head Start program operations;

(B) Head Start program quality;

(C) the length of time a child attends a Head Start program;

(D) the age of the child on entering the Head Start program;

(E) the type of organization (such as a local educational agency or a community action agency) providing services for the Head Start program;

(F) the number of hours and days of program operation of the Head Start program (such as whether the program is a full-working-day, full calendar year program, a part-day program, or a part-year program); and

(G) other characteristics and features of the Head Start program (such as geographic location, location in an urban or a rural service area, or participant characteristics), as appropriate.

(7) REPORTS.--

(A) SUBMISSION OF INTERIM REPORTS.--The organization shall prepare and submit to the Secretary two interim reports on the research. The first interim report shall describe the design of the research, and the rationale for the design, including a description of how potential sources of variation in impact of Head Start programs have been considered in designing the research. The second interim report shall describe the status of the study and preliminary findings of the study, as appropriate.

(B) SUBMISSION OF FINAL REPORT.--The organization shall prepare and submit to the Secretary a final report containing the findings of the research.

(C) TRANSMITTAL OF REPORTS TO CONGRESS.--

(i) IN GENERAL.--The Secretary shall transmit, to the committees described in clause (ii), the first interim report by September 30, 1999, the second interim report by September 30, 2001, and the final report by September 30, 2003.

(ii) COMMITTEES.--The committees referred to in clause (i) are the Committee on Education and the Workforce of the House of Representatives and the Committee on Labor and Human Resources of the Senate.

(8) DEFINITION.--In this subsection, the term 'impact', used with respect to a Head Start program, means a difference in an outcome for a participant in a program that would not have occurred without the participation in the program.

Appendix 1.2: Calculating Analytical Sampling Weights for Fall 2002 and Spring 2003

Overview

Sampling weights were calculated for each child and parent to allow estim ates based on the sample to represent the population of newly entering Head Start participants. Because children were randomly assigned to Head Start and non-Head Start groups within each Head Start center, each group represents the same Head Start population of newly entering children when appropriately weighted. The only difference, theoretically, is that the Head Start group was assigned to attend Head Start at the time of random assignment, while the non-Head Start group was not. Children who were sampled as Head Start group members or non-Head Start group members were assigned base weights that reflected their overall probability of selection, including the sampling of broad geographic areas used as primary sampling units (PSUs), Head Start grantees/delegate agencies, and centers. These base weights were adjusted for omission of programs and centers in communities saturated by Head Start and nonresponse to the fall 2002 and spring 2003 child assessment and parent interview separately to produce fall 2002 and spring 2003 child and parent weights, respectively. The nonresponse-adjusted weights of children in the 4-year-old group were poststratified to the Head Start National Reporting System (HSNRS) newly entering enrollment totals for 4-year-olds (comparable totals for 3-year-olds were not available). Extremely large weights were then trimmed for both age groups. The final child and parent weights are the product of the overall base weight, a nonresponse adjustment factor, a poststratification factor, and a trimming factor. For variance estimation, a set of 76 jackknife replicate weights was created for each child and parent.

Spring 2003 weights are used for most analyses in this report; the analyses focus on impacts at that time and include only children and families for whom spring data are available. Fall 2002 weights are used to examine distributions of child and family characteristics at the beginning of the analysis period, in fall 2002.

Primary Sampling Unit (PSU) Weights

The frame of 161 PSUs, or geographic clusters, was classified into 25 approximately equal-sized strata based on the level of services for low-income preschool children in the state, percentage of minority Head Start enrollment in the PSU, Head Start region, and percentage of Head Start enrollment in an MSA (a U.S. Census Bureau metropolitan statistical area). One PSU in each stratum was sampled with probability proportional to the total Head Start enrollment of 3- and 4-year-olds in the PSU. The source of enrollment was the 1999-2000 PIR. The PSU weight is the inverse of the PSU probability of selection:

> PSU weight = (Total Age 3 & 4 Enrollment in Stratum h) / (Total Age 3 & 4 Enrollment in PSU) where h = 1, 2, ….25. There was one certainty PSU whose probability of selection was 1 due to its large Head Start enrollment.

Head Start Program Weights

Program Sampling

There were two stages of sampling within most PSUs, and three stages within three extremely large PSUs. Prior to sampling, small programs were collapsed into groups consisting of two to four programs. These were sampled as a unit; thus, the within-PSU probability of selection for each program in a given group is the same.

Prior to telephone screening, programs and program groups (referred to henceforth simply as program groups, although most "groups" consisted of a single grantee or delegate agency) were sampled within the three large PSUs to reduce screening costs. In each of these three PSUs, 12 program groups were sampled with probability proportional to total age 3 and 4 enrollment from the 1999-2000 PIR. All programs in the sample PSUs underwent screening, during which study staff collected information on additional characteristics of each program and its community (except in the three large PSUs, where only the 12 sampled program groups were screened). A major purpose of this screening was to identify situations in which Head Start "saturated" the community, i.e., where the local program was large enough that all of the interested and eligible families in the community could be enrolled, making selection of a non-Head Start study group impossible without simultaneously leaving some of the program's capacity unused. After screening, program groups were sampled within the 25 PSUs from among those determined to be neither "saturated" nor closed. Within each PSU, four program groups were sampled with probability proportional to the total newly entering children ages 3 and 4 enrollment. From these, three program groups were subsampled with equal probabilities to be the main sample, and the remaining program group was assigned as a reserve sample. The main sample consisted of 76 program groups, which comprised 90 individual programs. The reserve sample consisted of 30 programs.

Program Base Weights, Adjustments for Saturation, Raking

Each of the 90 programs in the main sample received a base weight. The program base weight is the inverse of the overall probability of selection for the program, including the PSU probability of selection and the sampling of program groups within the PSU.

The base weights were adjusted for undercoverage due to the deletion from the frame of eight Head Start programs involved in the most recent FACES study and 28 programs discovered to be "saturated" during the screening. Because these programs had no chance of selection, an undercoverage adjustment was needed to correct for bias, in case the deleted programs were systematically different from those retained on the frame (see Appendix 2.1 for an examination of this question) and to prevent weighted enrollment totals from the sample from being too low. The undercoverage adjustment factor was calculated as the ratio of the estimated total newly entering enrollment in the PSU to the estimated newly entering enrollment from the sampled programs in the PSU, using enrollment information collected during the telephone screening. This adjustment corrected for differences between saturated and non-saturated programs on broad geographic factors but not for differences between the two types of programs within PSUs—differences that could result in larger or smaller Head Start impacts in the studied sites than in the nation as a whole.

The adjusted program weights for all 90 main sample programs were raked to marginal ages 3 and 4 enrollment totals from the 1999-2000 PIR. The raking dimensions were urban status (central city, noncentral city, rural), Head Start region (Northeast, North Central, South, Plains, West), and level of pre-K services in the state (state has Head Start-like programs, state has other types of programs, state has no programs). This procedure served to further match the analysis sample to the full national Head Start program on these factors. Since the number of sampled programs in each cross-classification is generally small, raking, or iterative proportional fitting (Oh & Scheuren, 1987), rather than poststratification was used. In raking, the weights are consecutively ratio-adjusted to marginal non-Head Start totals until the resulting weighted totals converge to the non-Head Start totals for each dimension. The adjustment factor at each iteration is the ratio of the PIR non-Head Start total for the marginal dimension to the sample estimate of the same total, where the weight in the sample estimate is the program weight from the previous raking iteration. This ratio adjustment reduces the sampling error associated with the sampling of PSUs and programs for estimates of Head Start children by urban status and Head Start region

(Cochran, 1977). However, it is not intended to result in sample estimates that will agree with non-Head Start totals of newly enrolled Head Start children, since no such counts exist.

After these undercoverage and raking adjustments were performed, the program weights in two PSUs were further adjusted to compensate for dropping two eligible programs from the sample because of their participation in a QRC study and for dropping three programs because they were found to be saturated after sampling. Another program was discovered to have closed, reducing the number of participating programs to 84. The adjustment factor was calculated as the ratio of estimated total newly entering enrollment in the PSU based on the entire sample of programs in the PSU to the weighted newly entering enrollment for the sampled nonsaturated, non-QRC programs in the PSU. None of the programs refused to participate, thus no nonresponse adjustment or reserve programs were needed.

Final Program Weight

Eighty-four programs received a final program weight. The final program weight can be written as:

Final program weight = PSU weight x $(1/ P_1)$ x $(1/ (1-P_{FACES}))$ x $(1/ P_2)$ x $(1/ P_3)$ x F_{Sat1} x F_{RK} x $F_{QRC, Sat2}$

where,

P_{FACES} = probability of selection in FACES,
P_1 = probability of being subsampled prior to telephone screening in three large PSUs,
P_2 = probability of being sampled in PSU,
P_3 = probability of being subsampled for main sample,
F_{Sat1} = adjustment factor for dropping 28 saturated programs from frame before sampling,
F_{RK} = raking adjustment factor to reduce sampling error,
$F_{QRC, Sat2}$= adjustment factor for dropping two programs participating in QRC and three saturated programs from the sample,

where,

P_1 = 12*(Total Age 3 & 4 Enrollment in Program Group)/(Total Age 3 & 4 Enrollment in PSU),
P_2 = 4*(1st Yr Age 3 & 4 Enrollment in Program Group)/(1st Yr Age 3 & 4 Enrollment in PSU),

The final program weights for the sample of 84 programs sum to 1,216 with a 95% confidence interval of [959, 1,472].

Head Start Centers

Center Sampling

Within each program, a list of the centers was obtained, and the centers were screened using a Center Information Form to collect various statistical data. The centers that were determined to be "saturated" were dropped from the frame in each program. Prior to sampling, small centers were combined into groups that ranged from two to eight centers and were treated as a unit for sampling purposes. Therefore, each center in a given group has the same probability of selection, namely that of the group. An initial sample of center groups was selected with probability proportional to newly entering age 3 and 4 enrollment in the center group. The initial sample of center groups was then subsampled with equal probabilities. The subsample was retained as the main sample in each program, while the remaining center groups formed a reserve sample. In general, three center groups per program (or program group) were selected for the main sample and two for the reserve. However, in very large programs four to six center groups were allocated for the main sample and three for the reserve. Within a program group, the total number of centers was allocated proportionally to the programs based on their newly entering enrollments. A total of 448 main sample and 237 reserve centers were selected in this way.

Center Base Weights and Adjustments for Saturation and Nonresponse

The center base weight is calculated as the inverse of the overall probability of selection for each center, including the sampling of PSUs, programs, and centers within programs. The center base weights were adjusted for deleting 161 saturated centers and 2 centers participating in a QRC study from the frame prior to center sampling. These adjusted weights were further adjusted for the refusal of 5 sampled centers to participate in the study, and for the loss of 56 centers discovered to be saturated after sampling. In these centers, no sampling of children was possible. In addition, 6 centers had closed, and 13 were ineligible for other reasons, such as merging with another center. For the merged centers, where appropriate, an adjustment was made to the base weight of the newly merged center to account for its increased probability of selection, since the individual centers had been listed separately on the center frame.

The adjustment factor for dropping saturated centers from the frame was calculated as the ratio of the estimated total newly entering enrollment in the program to the newly entering enrollment estimated from the sampled centers in the program. The newly entering enrollment was collected on the Center Information Form during center screening and updated during

October through December 2002 for all centers where possible. The adjustment factor was calculated separately for each program, unless this resulted in a very large adjustment, in which case the factor was calculated for the PSU.

The adjustment factor for the loss of five refusing and 56 saturated centers was calculated as the ratio of the weighted newly entering enrollment for the entire center sample in the program (excluding those that had closed or merged) to the weighted newly entering enrollment for the nonsaturated, cooperating centers in the program. Overall, these procedures adjusted for differences between included and excluded centers that emanate from the particular grantee or delegate agency that runs the excluded centers but not for other differences across centers that might lead to different-sized impacts in the omitted sites.

Final Center Weight

The final center weight can be written as:

$$\text{Final Center Weight} = \text{Final Program Weight} \times (1/P_{c1}) \times (1/P_{c2}) \times F_{QRC} \times F_{Sat1} \times F_{Refusal,\ Sat2},$$

where,

P_{C1} = probability of selection for initial center sample (both main and reserve),
P_{C2} = probability of selection for main center sample,
F_{QRC} = adjustment factor for dropping two centers participating in QRC from frame,
F_{Sat1} = adjustment factor for dropping 161 saturated centers from frame,
$F_{Refusal,\ Sat2}$ = adjustment factor for dropping 56 saturated centers and 5 refusing centers from sample,

$$P_{C1} = \frac{\text{Newly Entering Age 3 \& 4 Enrollment in Center Group}}{(\text{Newly Entering Age 3 \& 4 Enrollment in Program for Eligible, Nonsaturated Centers})/n_{M+R}},$$

$$P_{C2} = \frac{n_M}{n_{M+R}} = \frac{\text{\# center groups subsampled for main sample in the program}}{\text{\# center groups sampled for both main, reserve in the program}},$$

and the final program weight reflects the PSU and program probabilities of selection. In four programs, all reserve centers were brought into the sample when the original centers were found to be saturated or partially saturated and hence unable to provide the planned number of non-Head Start sample children. In these centers, P_{C2} was set to one in the above formula. When this resulted in a census of eligible centers in the program, both P_{c1} and P_{c2} were set to one. In six programs where some, but not all, of the reserve centers were activated to offset saturation in the main sample, n_M includes the reserves that were activated as well as the main sample centers.

In this situation, centers were randomly subsampled from among the reserve centers selected for that particular program or program group. The total number of centers in the final sample, including main sample and activated reserves is 458. The sample was reduced to 378 after losing 19 centers identified following selection as ineligible (closings, mergers), 5 identified as noncooperating, and 56 found to be saturated.

Because reserve centers were picked at random from the same pool as the main sample centers, utilization of the reserve sample will bias study results only to the extent that the centers they replaced were atypical. Hence, recourse to reserve sampling represents another part of the study's overall undercoverage of communities saturated by Head Start.

The final center weights for the 378 centers sum to 12,705 with a 95% confidence interval of [10,290, 15,119].

Child Weights

Random Assignment of Children Within Centers

Children were sampled in two stages within each center. At the first stage, the applicant list was sorted based on child need, and the list was truncated at exactly the number of children needed to both fill the center's slots and supply a non-Head Start group sample of the desired size for the study. A sample of children was then randomly selected with equal probabilities from the truncated list to fill the center's slots. Those not selected to fill a slot were assigned to the non-Head Start group. At the second stage, the children sampled to fill the center's slots were subsampled to obtain the targeted number of Head Start group children. Thus, there were four categories of children: 1) those sampled to attend the Head Start program but not for participation in the study, 2) those sampled for the study's Head Start group, 3) those sampled for the study's non-Head Start group, and 4) those on the waiting list who had no chance of selection for either study sample but who could enter the Head Start program later (once sampling ended) to replace children who dropped out of the program over the course of a year. The targeted number of Head Start and non-Head Start group children was 16 and 11, respectively, at most centers and center groups, cumulating to an average of 48 Head Start group members and 32 non-Head Start group cases for each sampled program group. In center groups, the 16 Head Start and 11 non-Head Start were proportionally allocated to the centers in the group based on newly entering enrollment. In 3 of the 84 programs, children applied directly to the program rather than the center, so it was necessary to randomly assign children at the program level and sample 48 Head

Start and 32 non-Head Start cases to obtain 80 children for the program in total. The total target sample size was approximately 3,600 Head Start and 2,400 non-Head Start children.

The random assignment of children was spread out over the summer/fall 2002, because most centers took applicants on a flow basis and preferred to let their families know soon whether their child had been accepted to attend the Head Start program. This meant children were sampled in batches or rounds, and the two-stage sampling process described above took place more than once in most centers. An additional complication was that stratification by program option was used in many centers. The allocation of the total number of Head Start and non-Head Start children across program options and rounds at each center was approximately proportional to the newly entering enrollment in each program option and the number of slots filled in each round. The actual probabilities of selection for each child were stored electronically for weighting purposes. However, the probabilities can vary greatly because of the difficulty in allocating across rounds. There were many rounds where children were sampled to fill slots but no Head Start or non-Head Start children were selected because the target sample sizes of Head Start and non-Head Start children had already been obtained. None of these children had a chance of selection for the study, meaning child weights based on the actual probabilities of selection would underestimate the size of the first year Head Start population.

Child Base Weights

The within-center child base weight was calculated as:

$$\frac{\text{Newly Entering Age 3 \& 4 Enrollment in Center}}{\text{\# treatment children sampled in center}}$$

for the sampled Head Start group children, and as

$$\frac{\text{Newly Entering Age 3 \& 4 Enrollment in Center}}{\text{\# control children sampled in center}}$$

for the non-Head Start group children. Note that the numerator is the same for both groups, since estimates are to be made for the universe of newly entering Head Start children using either sample. For centers where the updated fall 2002 newly entering enrollment was not obtained, the newly entering enrollment figure for the previous program year was used. When this was missing, and for three programs where children were randomly assigned at the program level rather than at the center level, the inverse of the actual probability of selection for children in the center was used as the base weight.

The overall child base weight reflecting all stages of sampling can be written as:

Overall Child Base Weight = (Final Center Wt) x (Within-Center Child Base Wt.)

where the final center weight reflects the PSU and program probabilities of selection and includes an adjustment for centers where no children were sampled because of center noncooperation or saturation.

Nonresponse Adjustments

Nonresponse adjustments were performed separately for fall 2002 and spring 2003, using three definitions of a respondent for the fall 2002 data collection and two definitions for spring 2003. The three definitions for fall 2002 were (1) child is considered a complete for the fall 2002 child assessment, (2) child has a complete fall 2002 parent interview, and (3) child is considered a complete for both the fall 2002 child assessment and parent interview. The two definitions for spring 2003 were (1) child is considered complete for the spring 2003 child assessment and (2) child has a completed spring 2003 parent interview. This resulted in three nonresponse-adjusted child weights for fall 2002 and two for spring 2003.

The nonresponse adjustment helps non-Head Start nonresponse bias by compensating for different data collection response rates across various demographic and geographic groups of children. This is due to the fact that the nonresponse adjustment factor is calculated within nonresponse adjustment cells formed by the demographic and geographic variables. The nonresponse adjustment factor spreads the weight of the nonresponding children over the responding children in that cell, so that they represent not only children who were not sampled, but also the nonresponding sampled children. This maintains the same mix of the sample across cells as would have been present had there been no nonresponse.

To capture the variation in response rates, we form cells based on characteristics that correlate with response rates. For the fall 2002 nonresponse adjustments, a nonresponse analysis using chi-square tests and logistic regression in WesVar showed high correlation between response rates and Head Start versus non-Head Start assignment and program option for the non-Head Starts. This result, combined with a desire to capture individual Head Start program differences as much as possible, led to nonresponse adjustment cells formed by crossing PSU x state x program for the Head Start group, and PSU x program option x state x program for the non-Head Start group. Collapsing across program and state was done as needed to prevent weight adjustment factors of 2.0 or larger.

To determine the nonresponse adjustment cells for spring 2003, an unweighted nonresponse analysis was done using a software package called CHAID (Chi-squared Automatic Interaction Detector), to determine what variables are correlated with propensity to respond. The following variables were used as candidates in the analysis:

- Head Start versus non-Head Start group,
- Child race,
- Child language,
- Language spoken at home,
- Child's gender,
- Program option applied for (full-day, part-day, both, home-based),
- Child's age,
- Metro status for county containing Head Start program office,
- Level of pre-K services in the state,
- Head Start region,
- State,
- Response status for fall 2002 child assessment,
- Response status for fall 2002 parent interview,
- Program, and
- PSU.

A small number of missing values for the variables used in the nonresponse analysis were imputed via hot deck imputation using procedures described in Appendix 4.1. Variables with missing values were child language, home language, child race, and gender. Weighted logistic regression and chi-square tests were also run in WesVar to confirm the CHAID results.

The tree structure identified by CHAID was used in creating the nonresponse adjustment cells for spring 2003. For the child assessment nonresponse adjustment, CHAID used the following variables to create nonresponse adjustment cells:

- Head Start versus non-Head Start indicator,
- Fall 2002 child assessment response status,
- Level of pre-K services in state,

- PSU,
- Head Start region,
- Child's gender,
- Metro status, and
- Child's race.

For the parent interview, the nonresponse adjustment cells were created using:

- The Head Start versus non-Head Start indicator,
- Fall 2002 parent interview response status,
- Level of pre-K services in state,
- PSU,
- Head Start region,
- Child's gender,
- Metro status,
- Child's age, and
- Child's race.

Some collapsing of cells was required to prevent excessively large nonresponse adjustment factors, which cause the weights to become more variable and the variance of most estimates from the data to increase. The coefficient of variation of the nonresponse-adjusted child weights was computed under various cell-collapsing scenarios for the child assessment and parent interview nonresponse adjustment for spring 2003. A final set of collapsed cells for each nonresponse adjustment was chosen based on a compromise between limiting the increase in weight variability and the need to control for non-Head Start for nonresponse bias by limiting the amount of cell collapsing.

Poststratification

To reduce the sampling error for estimates of the newly entering Head Start population, the nonresponse-adjusted child weights for children in the 4-year-old group were poststratified to fall 2003 HSNRS newly entering enrollment totals by race/ethnicity. (The HSNRS is a census of Head Start programs, so there should be no sampling error associated with its enrollment totals. However, race reporting may differ somewhat between the HSNRS and the current study, as the Head Start programs were given no specific instructions on how to code the variable in the HSNRS.) Comparable enrollment totals were not available for 3-year-olds. The three race/ethnicity categories were Hispanic, non-Hispanic, Black, and White/other. An adjustment factor was calculated for each category, and the appropriate factor applied to each child weight depending on the race of the child, as reported on the NHIS child roster. The numerator of each factor was the proportion of HSNRS total newly entering age 4 enrollment in the race/ethnicity category; the denominator was the sample estimate of this proportion using the 84 programs sampled for the current study, the final program weight, and the HSNRS first year age 4 enrollment reported for each program. The poststratification factors were 0.80 for Hispanic, 1.45 for Black, and 1.036 for White/other, indicating an overrepresentation of Hispanic children and underrepresentation of Black children in the current study sample as compared to the HSNRS. Appendix 2.3 provides a detailed analysis of the race/ethnicity composition of the sample and its comparison to national Head Start data.

Trimming

A final trimming adjustment was made for inordinately large child weights. Very large weights can substantially increase sampling error, so weights were trimmed back to four times the average weight to avoid large sampling errors, even though this introduces a small amount of bias into the survey estimates. For the fall 2002 child weights, 76 weights (2.0%) were trimmed for the child assessment completes, 79 (2.0%) for the parent interview completes, and 75 (2.0%) for children having both a complete child assessment and parent interview. For the spring 2003 child weights, 84 weights (2.2%) were trimmed for the child assessment completes and 86 (2.2%) for the parent interview completes. An analysis of the trimmed cases showed that most extremely large weights were primarily due to some large centers being undersampled, i.e., only a few children were sampled, perhaps due to near-saturation.

The final child weight can be written as:

Final Child Weight = (Overall Child Base Wt) x (Child Nonresponse Adjustment Factor)
x (Poststratification Factor) x (Trimming Factor)

where the overall child base weight reflects the probability of selecting the PSU, program, center, and child within center. When the final child weight is applied, the Head Start and non-Head Start groups each separately represent the entire first year Head Start population. Sample estimates of the size of the first year Head Start population are given in Exhibit A.1.2.1 in the "Sum of Final Weights" column.

Exhibit A.1.2.1: Final Sampling Weights, Fall 2002 and Spring 2003

	Number of Respondents	Sum of Final Weights	95 Percent Confidence Interval	Coefficient of Variation of Final Weights
Final Fall 2002 Child Weights				
Child Assessment				
Head Start	2,360	422,686	(352,936, 492,437)	0.860
Non-Head Start	1,363	413,258	(345,160, 481,356)	0.770
Parent Interview				
Head Start	2,489	423,086	(353,623, 492,548)	0.850
Non-Head Start	1,526	414,214	(346,413, 482,016)	0.780
Both Child Assessment and Parent Interview				
Head Start	2,339	422,818	(353,030, 492,606)	0.860
Non-Head Start	1,361	413,064	(345,221, 480,907)	0.770
Final Spring 2003 Child Weights				
Child Assessment				
Head Start	2,441	426,834	(357,492, 496,177)	0.860
Non-Head Start	1,457	418,907	(352,648, 485,166)	0.880
Parent Interview				
Head Start	2,404	427,536	(358,628, 496,444)	0.860
Non-Head Start	1,483	419,772	(353,164, 486,381)	0.880

Reweighting Non-Head Start Group Observations After Deleting Crossovers

A crossover is defined as a child who was randomly assigned to the non-Head Start group but participated in Head Start. Of the 227 crossovers in the sample, 212 were respondents for the spring 2003 child assessments, and 211 had a completed parent interview in spring 2003.[1] To develop alternative "crossover-adjusted" estimates of Head Start's impact to supplement the main findings, these cases were dropped from the analysis sample, and weights for the remaining non-Head Start group members were recalculated. In effect, this procedure treated crossovers as a second set of nonrespondents to the spring 2003 data collection.

This additional nonresponse adjustment took as its starting point the previously nonresponse-adjusted child assessment and parent interview spring 2003 child weights. It then inserted an additional stage of nonresponse adjustment for the crossovers just prior to the poststratification to the HSNRS totals. A CHAID analysis was run using demographic characteristics of the child and parents, household income variables, and health-related questions from the parent interview as inputs. A minimum cell size of 30 was required and a minimum p-value of 0.05 (with a Bonferroni adjustment) was required for retention in the tree. At the top of the tree, age group (age 3 or 4) was forced to be the first variable because the cohorts were analyzed separately and because a logistic regression analysis of crossover patterns indicated a significant age by gender interaction.

For the 3-year-old group, CHAID identified five groupings of PSUs with similar unweighted crossover rates. It then split one of these groupings by father's immigration status, another by parent-reported emergent literacy scale for the child in fall 2002 and food stamp receipt, a third by mother's employment status in the fall, and a fourth by teen birth status of the mother—creating 10 cells in total. For the 4-year-old group, CHAID identified four groupings of PSUs with similar unweighted crossover rates. It then split one of these groupings by child's gender, creating a total of 5 cells. (No other correlates with crossover rate were identified for the remaining PSU groups.)

A crossover "nonresponse" adjustment factor was then calculated for each cell to spread the weight of deleted crossover cases over the remaining non-crossover observations in that cell, so that the latter could represent crossover-like children in a "non-treated" state. For each non-crossover non-Head Start in the cell, the crossover adjustment factor was multiplied by the pre-existing nonresponse-adjusted weight for that person. The resulting weights were then poststratified and trimmed as before. Separate crossover nonresponse adjustments were done in this manner for spring 2003 child assessment outcomes and spring 2003 parent interview outcomes.

Analysis weights for randomly assigned Head Start group children remained unchanged when conducting the analysis of crossover-adjusted impacts.

Importance of Using Weights

The formulas for producing weights are quite complex and can result in substantial differences in weights among sample children. If certain types of children tend to have much larger weights than other types of children, and if the weights are not used in the analysis, then the types of children with large weights will be underrepresented in the analysis relative to the population of all newly entering Head Start children. This can lead to serious bias in impact estimates. Thus, we strongly recommend that weights be used in all analyses.

---

[1] The overall weighted crossover rate for the non-Head Start group was 17.6 percent.

Calculating Correct Standard Errors

Estimates obtained from the Head Start Impact Study will differ from the true population parameters because they are based on a randomly chosen subset of the population, rather than on a complete census of all newly entering Head Start children. This type of error is known as **sampling error** or **variance**. The differences between the estimates and the true population values can also be caused by **nonsampling error**. Nonsampling errors can result from many causes, such as measurement error, nonresponse, sampling frame errors, respondent error, and differences among interviewers. In general, the magnitude of nonsampling error is difficult to assess from the sample. The **precision** of an estimate is measured by the standard error (defined as the square root of the variance). The calculation of the standard error must reflect not only the sample size on which the estimate is based, but the manner in which the sample was drawn. Otherwise, the standard errors can be misleading and result in incorrect confidence intervals and p-values in hypothesis testing. The study's sampling involved stratification, clustering, and unequal probabilities of selection, all of which must be reflected in the standard error calculations.

Two commonly used variance estimation methods for complex surveys involving multi-stage sampling are replication and linearization (Wolter, 1985). Replication methods work by dividing the sample into subsample replicates that mirror the design of the sample. A weight is calculated for each replicate using the same procedures as for the full-sample weight. This produces a set of replicate weights for each sampled child. To calculate the standard error of a survey estimate, the estimate is first calculated for each replicate using the replicate weight and the same form of estimator as for the full sample. The variation among the replicates is then used to estimate the variance for the full sample estimate. In the linearization approach, a nonlinear estimator is approximated by a linear function and a formula derived for the variance of the linear approximation. Replication has the advantage that it can reflect the different features of the weighting and estimation by simply repeating all steps separately for each replicate. For linearization, a specific formula is needed for each estimator, and the formula will differ depending on the type of estimator and sample design. On the other hand, finite population correction factors are often easier to account for using linearization estimators. However, for linear estimators, or nonlinear estimators that are formed by combinations of linear functions, replication variance estimators are often little different numerically from linearization variance estimators.

For the current study, a set of jackknife replicate weights was created for each child for use in the calculation of standard errors. Normally, stratified jackknife replicate weights are created by dropping out one PSU at a time, setting the replicate weights for sampled units in the dropped PSU to zero, multiplying the full-sample weights of sampled units in the remaining PSUs in the stratum by a factor of n_h / (n_h-1), where n_h is the number of PSUs in the h-th stratum, and leaving the full-sample weights for sampled units in the remaining strata unchanged. However, because only 25 PSUs were sampled at the first stage (one per stratum), only 27 replicate weights could be created (in the one certainty PSU, two additional replicates could be formed based on program groups). To improve the stability of the variance estimates, the second-stage sampling units, namely Head Start program groups, were used as the “drop unit” in creating replicates. This resulted in 76 replicate weights per child and 51 degrees of freedom for variance estimation (i.e., 76 PSUs – 25 strata). Because the between-PSU component of variance is being ignored in doing this, the resulting variance estimates will be slight underestimates, if the between-PSU variability is small relative to the within-PSU variability.

The validity of this hypothesis was investigated by creating a second set of 27 replicate weights based on the 25 PSUs, which includes the between-PSU component, but has fewer degrees of freedom. By calculating the average ratio of the variance from the set of replicate weights based on the 25 PSUs to the variance from the set based on the 76 program groups, we were able to estimate the relative size of the between-PSU component. The ratio of variances was calculated for several child assessment means (PPVT, Elision, Woodcock-Johnson Applied, Oral Comprehension, Spelling, and Letter-Word) by age and gender within the test language groups English and Spanish and averaged them. (However spring 2003 variance estimates could not be produced separately for the Spanish group because the completed Spanish assessments were all from only three sampled Head Start programs, resulting in insufficient degrees of freedom to estimate the variance.) For fall 2002 scores, the between-PSU component was estimated to be 15 percent of the total variance, and for spring 2003 scores, this component was estimated to be 28 percent of the total variance. Therefore, estimates of fall 2002 standard errors from the fall 2002 replicate weights should be multiplied by the square root of 1.15 (=1.07) to prevent underestimates of the variance. Similarly, standard errors for spring 2003 based on the spring 2003 76 replicates should be multiplied by the square root of 1.28 (=1.13).

Incorporating Weights and Standard Errors in the Impact Analyses

The easiest way for analysts to incorporate the weights and correct standard errors into their analyses will be to use software designed for analysis of complex survey data. Such software packages include Wesvar, SUDAAN, Stata, and the new survey procedures (proc surveymeans, proc surveyreg) in SAS version 8. SAS version 9 will add a logistic regression procedure for survey data. Most estimation and modeling can be done with one of these packages, with the possible exception of hierarchical linear modeling (HLM). WesVar uses replication methods (jackknife, BRR), and Stata and SAS version 8 use linearization. SUDAAN uses both linearization and replication.

REFERENCES

Cochran, W. G. (1977). ***Sampling Techniques.*** New York: Wiley & Sons, ch. 6.

Oh, H.L. & F.J. Scheuren. (1987). "Modified raking ratio estimation." ***Survey Methodology*, 13,** 209-219.

Research Triangle Institute. (2001). ***SUDAAN User's Manual, Release 8.0.*** Research Triangle Park, NC: Author.

StataCorp. (2001). ***Stata Statistical Software: Release 7.0***. College Station, TX: Author.

Wolter, K. (1985). ***Introduction to Variance Estimation***. New York: Springer-Verlag.

WesVar (2003). ***WesVar™ 4.2 User's Guide.*** Rockville, MD: Westat.

Appendix 1.3: Language Decision Form

To the best of your knowledge,

1. What language does the child speak most often at home?

   ENGLISH 01
   SPANISH 02
   OTHER (SPECIFY) 03
   ____

2. What language does the child speak most often at this child care setting?

   ENGLISH 01
   SPANISH 02
   OTHER (SPECIFY) 03
   ____

3. What language does it appear this child prefers to speak?

   ENGLISH 01
   SPANISH 02
   OTHER (SPECIFY) 03
   ____

Language in which at least two of three responses are the same:

____
LANGUAGE

4. If language is other than English or Spanish, ask main care provider: Can child understand and answer questions in English? (IF YES, PROCEED WITH ENGLISH TESTING. OTHERWISE FOLLOW INSTRUCTIONS FOR CHILDREN BEING TESTED IN OTHER LANGUAGE)

   YES 1
   NO 2

5. Language child will be tested in:

____
LANGUAGE

Appendix 1.4: Citations for Child Assessments, Scales, and Observation Instruments

CHILD ASSESSMENT BATTERY

Dunn, L.M., L.L. Dunn, and D.M. Dunn. (1997). ***Peabody Picture and Vocabulary Test, Third Edition (PPVT)***. Circle Pines, MN: American Guidance Service.

Dunn, L.M., E.R. Padilla, D.E. Lugo, and L.L. Dunn. (1986). ***Test de Vocabulario en Imágenes Peabody.*** Circle Pines, MN: American Guidance Service.

FACES Research Team. Color Names and Counting. Modified from the Color Concepts and Number Concepts Tasks in J.M. Mason and J.Stewart. (1989). ***The CAP Early Childhood Diagnostic Instrument (prepublication edition),*** American Testronics.

FACES Research Team. Letter Naming Task. Modified from a test used in the Head Start Quality Research Center's curricular intervention studies.

FACES Research Team. Story and Print Concepts. Modified from the Story and Print Concepts in J.M. Mason and J. Stewart. (1989). ***The CAP Early Childhood Diagnostic Instrument (prepublication edition),*** American Testronics.

FACES Research Team. Writing Sample. Modified from the Name Writing Tasks in J.M. Mason and J. Stewart. (1989). ***The CAP Early Childhood Diagnostic Instrument (prepublication edition),*** American Testronics.

Leiter-R AM Battery. (1997). Wood Dale, IL: Stoelting Co. (Subtest: Attention Sustained).

Lonigan, C.J., R.K. Wagner, J.K. Torgesen, and C. Rashotte. (2002). ***Preschool Comprehensive Test of Phonological & Print Processing.*** (Subtests: Print Awareness and Elision).

McCarthy, D. (1970, 1972). ***McCarthy Scales of Children's Abilities***. San Antonio, TX: The Psychological Corporation. (Subtest: Draw-a-Design Task).

Woodcock, R.W., K.S. McGrew, and N. Mather. (2001). ***Woodcock-Johnson III Tests of Achievement***. Itasca, IL: Riverside Publishing. (Subtests: Letter-Word Recognition, Spelling, Oral Comprehension, and Applied Problems).

Woodcock, R.W. and A.F. Munoz-Sandoval. (1996). ***Batería Woodcock-Muñoz Pruebas de aprovechamiento-Revisada.*** Itasca, IL: Riverside Publishing. (Subtests: Identificaci⊠n de letras y palabras, Dictado, and Problemas aplicados).

TEACHER/CARE PROVIDER CHILD REPORT

High Scope Educational Research Foundation. (1992). ***Child Observation Record (COR).*** Ypsilanti, MI: Author.

Lutz, M.N., J.F. Fantuzzo, and P. McDermott. (in press). "Adjustment Scales for Preschool Intervention." ***Early Childhood Research Quarterly.***

Pianta, R.C. (1992). ***Student-Teacher Relationship Scale.*** Charlottesville, VA: University of Virginia.

QUALITY OF CARE OBSERVATIONS

Harms, T., R.M. Clifford, and D. Cryer. (1998). ***Early Childhood Environment Rating Scale-Revised Edition (ECERS-R).*** New York, NY: Teachers College Press.

Harms, T. and R.M. Clifford. (1989). ***Family Day Care Rating Scale (FDCRS)***. New York, NY: Teachers College Press.

Arnett, J. (1989). "Caregivers in day-care centers: Does training matter?" ***Journal of Applied Developmental Psychology***, ***10***, 541-552.

PARENT INTERVIEW SCALES

Achenbach, T.M., C. Edelbrock, and C.T. Howell. (1987). "Empirically Based Assessment of the Behavioral/Emotional Problems of 2-3 Year-Old Children." ***Journal of Abnormal Child Psychology, 15***, 629-650.

Developing Skills Checklist—Home Inventory. (1990). Monterey, CA: CTB/McGraw-Hill.

Entwistle, D.R., K.L. Alexander, D. Cadigan, and P.M. Pallis. (1987). "The Emergent Academic Self-Image of First Graders: Its Response to Social Structure. ***Child Development, 58***, 1190-1206.

Perlin, L.I. and C. Schooler. (1978). "The Structure of Coping." ***Journal of Health and Social Behavior*, *22,*** 337-356. (Pearlin Mastery Scale-Locus of Control).

Pianta, R.C. (1992). ***Parent-Child Relationship Scale.*** Charlottesville, VA: University of Virginia.

Ross, C.E., J. Mirowsky, and J. Huber. (1983). "Dividing Work, Sharing Work, and In-Between: Marriage Patterns and Depression." ***American Sociological Review*, *48,*** 809-823.

Appendix 2.1: Comparison of Head Start Grantees/Delegate Agencies and Centers in Saturated and Non-Saturated Communities

As discussed in Chapter 2, there is potential for undercoverage bias due to the exclusion from the sampling frame of Head Start grantees/delegate agencies and centers in communities saturated by the program, i.e., communities with too few extra families interested in Head Start (beyond those the program can accommodate) to provide a randomly selected non-Head Start group for the study. Newly entering Head Start children in these saturated communities had no chance of selection and therefore are not represented by our sample. Consequently, the potential for bias arises if the saturated grantees/delegate agencies and centers are systematically different from the non-saturated grantees/delegate agencies and centers we retained in the sampling frame and if the characteristics on which they differ are correlated with the outcome measures for and impact estimates on the children they enroll. However, if the children in these excluded grantees/delegate agencies and centers represent only a small percentage of the Head Start population, then the potential for bias is much less. Based on the sample coverage rate reported in Chapter 2, 15.5 percent of the children served by Head Start nationally are omitted from the study. This noncoverage rate is based on grantees and centers identified in the sample frame and samples that were excluded due to saturation. It equals 1 minus the product of four coverage rates: program frame x program sample x center frame x center sample. Mathematically, this equates to 1-(0.962 x 0.975 x 0.952 x 0.947) = 1-0.845 = 0.155.

Head Start Grantees/Delegate Agencies

Exhibits A.2.1.1 and A.2.1.2 compare saturated and non-saturated grantees/delegate agencies by a few qualitative characteristics and enrollment variables available on the Head Start Program Information Report (PIR) database (and, for newly entering enrollment, telephone screening confirmation calls to grantees and delegate agencies prior to sampling). The grantees/delegate agencies were weighted to account for sampling of broad geographic areas (i.e., PSUs) and for the subsampling of grantees/delegate agencies in three large urban cities prior to the telephone screening (see Chapter 1). This is necessary to draw conclusions about the entire population of children served by Head Start and not merely the children served by grantees/delegate agencies in the 25 sampled PSUs that were screened to determine saturation. Tests of statistical significance were performed to reduce the possibility of drawing false conclusions from differences that may be due to sampling error. The hypothesis testing was done in WesVar using jackknife replicate weights to account for the study's complex sample design.

As shown in these tables, the saturated grantees/delegate agencies are much smaller, much more likely to be school-based, and have smaller percentages of Hispanic enrollment than the non-saturated grantees/delegate agencies. Although they appear to be more often located in the midwest, differences in the distribution of saturated vs. non-saturated grantees/delegate agencies by Head Start regions are not statistically significant. A cautionary note is that variances at the program level are not very stable because the number of saturated grantees/delegate agencies is small. In addition, variances do not include the between-PSU component of variance due to sampling PSUs; thus, they are underestimates, and the p-values may be slightly overstating the significance of the differences.

Exhibit A.2.1.1: Comparison of Saturated and Non-Saturated Head Start Grantees/Delegate Agencies by Enrollment

Enrollment Variable	Saturated Programs	Non-Saturated Programs	P-Value (t-Test of Difference)
Percent Hispanic Enrollment	9%	26%	**0.001**
Percent Black Enrollment	20%	33%	0.134
Age 3 Enrollment as Percent of Total Enrollment	52%	49%	0.535
Average Total Enrollment	188	571	**<0.001**
Average Newly Entering Enrollment	113	388	**<0.001**

Exhibit A.2.1.2: Comparison of Saturated and Non-Saturated Head Start Grantees/Delegate Agencies by Location Characteristics

Characteristics	Saturated Programs	Non-Saturated Programs	P-Value (Chi-Square Test of Association)
School-based			**0.018**
Yes	66%	21%	
No	34%	79%	
Metro Status			0.91
MSA	66%	68%	
Non-MSA	34%	32%	
Level of Pre-K Services in State			0.60
Similar to Head Start	35%	25%	
Some Head Start-Like	27%	20%	
Remaining States	38%	55%	
Head Start Region			0.15
Northeast	24%	25%	
Midwest	48%	24%	
South	28%	39%	
Plains	0%	4%	
West	0%	8%	

Head Start Centers

Exhibits A.2.1.3 and A.2.1.4 compare saturated and non-saturated centers by various qualitative characteristics and enrollment variables available from Center Information Forms (CIFs) completed by all centers in the sampled grantees and delegate agencies. All hypothesis testing was again done in WesVar using jackknife replicate weights to account for the study sample design. The replicate weights do not include the between-PSU variance component, therefore the p-values in these tables may slightly overstate the significance of the difference. In Exhibit A.2.1.3 the chi-square test was not able to detect a significant difference for type of program option offered, whether staff are school employees, metro status, region, or level of Pre-K services available in the state. With respect to enrollment, Exhibit A.2.1.4 shows that the saturated centers are smaller, have fewer Hispanic children, and have a larger percentage of first year 3-year-olds than the non-saturated centers. As expected, these centers do not have waiting lists, a significant difference from non-saturated centers.

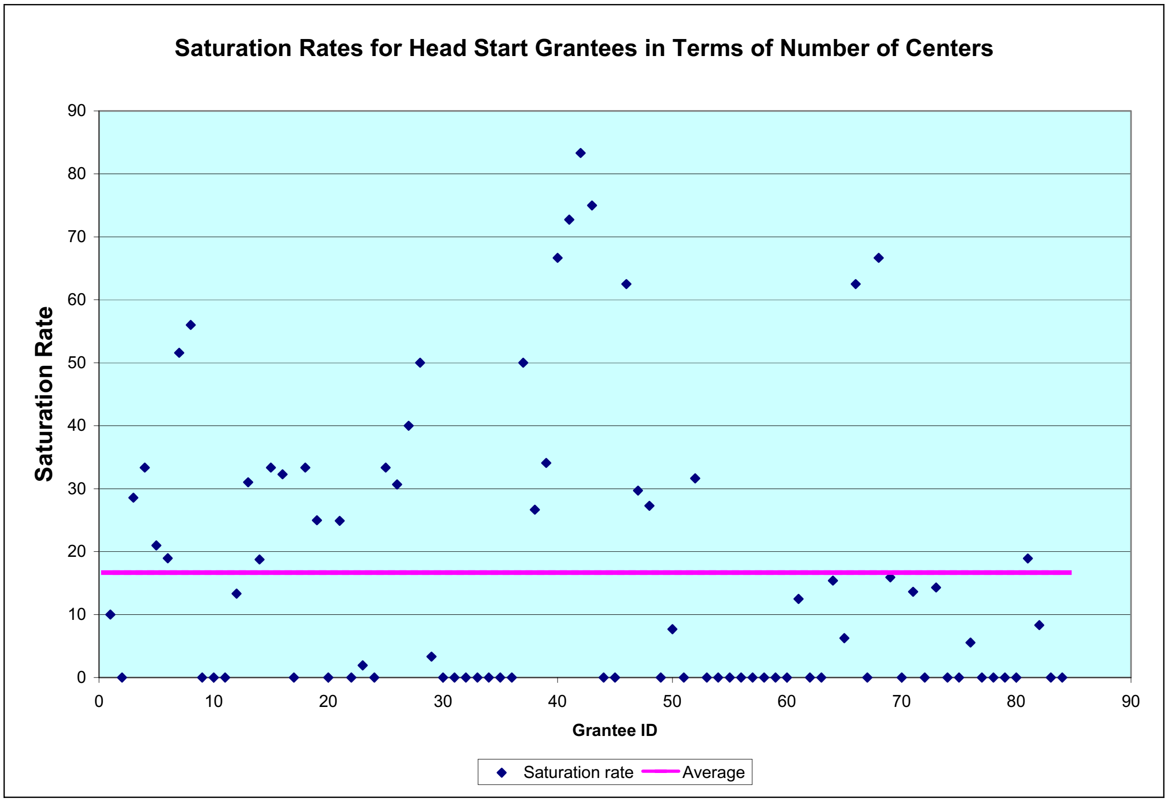
Two graphs follow Exhibit A.2.1.4 that show the percentage of centers that are saturated for each of the 84 grantees/delegate agencies with less than 100 percent saturation rate. The saturation rate was calculated two ways: as the percentage of centers in each program that are saturated and as the percentage of newly entering enrollment in saturated centers for each program. The average percentage of saturated centers is 16.6 percent while the average percentage of newly entering enrollment in saturated centers is 13.2 percent, another indication that the saturated centers tend to be smaller. The graphs show the extreme variation among grantees/delegate agencies in the share of centers operating in saturated communities and the share of newly entering children served by those centers.

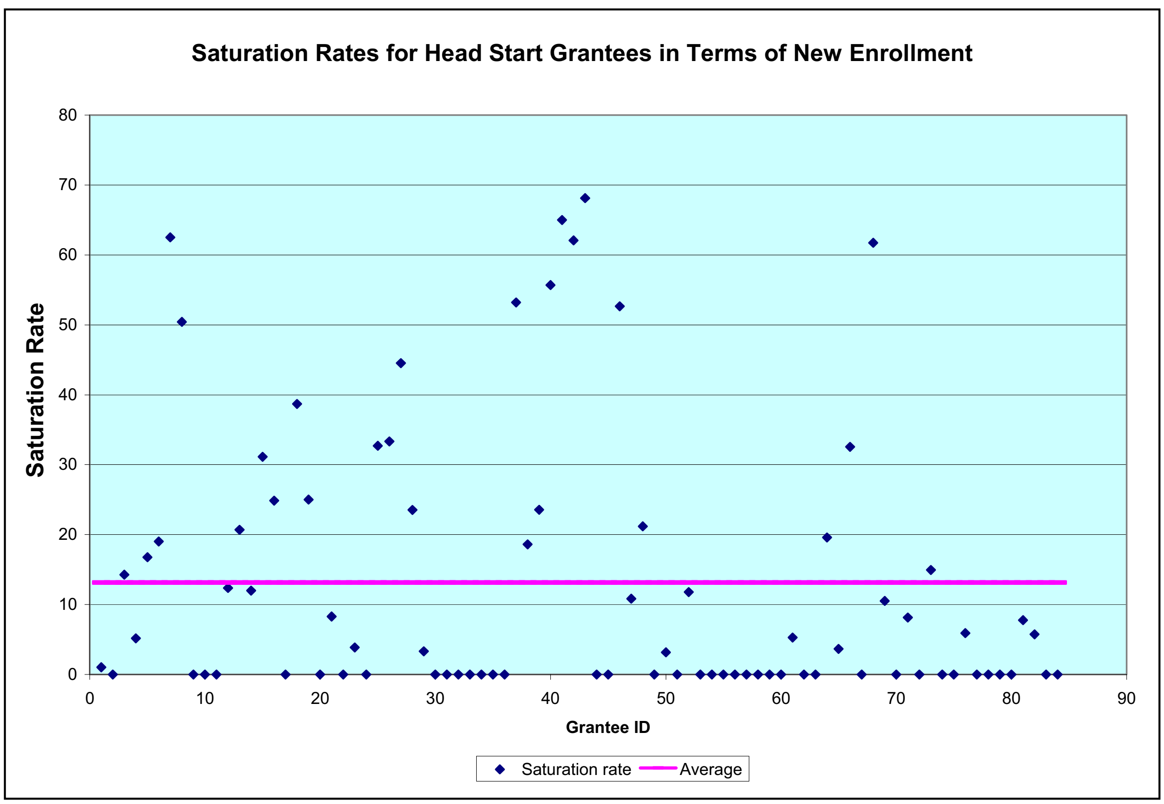
Exhibit A.2.1.3: Comparison of Saturated and Non-Saturated Head Start Centers Operated by Non-Saturated Programs, by Program and Location Characteristics

Characteristics	Saturated Centers	Non-Saturated Centers	P-Value (Chi-Square Test of Association)
Program Option			0.44
Full-Day Only	35%	28%	
Part-Day Only	52%	50%	
Other	13%	22%	
Staff Are School Employees			0.249
Yes	17%	11%	
No	83%	89%	
Metro Status			0.64
MSA	74%	70%	
Non-MSA	26%	30%	
Head Start Region			0.376
Northeast	32%	27%	
Midwest	34%	20%	
South	17%	31%	
Plains	12%	11%	
West	4%	11%	
Level of Pre-K Services in State			0.212
Similar to Head Start	40%	22%	
Some Head Start-Like	15%	18%	
Remaining States	45%	60%	

Exhibit A.2.1.4: Comparison of Saturated and Non-Saturated Head Start Centers Operated by Non-Saturated Programs, by Enrollment

Enrollment Characteristic	Saturated Centers	Non-Saturated Centers	P-Value (t-test of Difference)
Percent Hispanic Enrollment	17%	30%	**0.005**
Percent Black Enrollment	38%	26%	0.204
Percent Newly Entering Enrollment	65%	66%	0.985
Age 3 Enrollment as Percent of Newly Entering Enrollment	54%	47%	**0.037**
Number of Children on Waiting List as Percent of Total Enrollment	0%	15%	**<0.001**
Average Number Funded Slots	37	48	**0.036**
Average Total Enrollment	26	47	**<0.001**
Average Newly Entering Enrollment	16	31	**<0.001**
Average Number on Waiting List	0	9	**<0.001**


Saturation Rates for Head Start Grantees in Terms of Number of Centers
Saturation Rate
Grantee ID
0
10
20
30
40
50
60
70
80
90
Saturation rate
Average


Saturation Rates for Head Start Grantees in Terms of New Enrollment
Saturation Rate
Grantee ID
0
10
20
30
40
50
60
70
80
90
Saturation rate
Average

Appendix 2.2: Determination of Head Start Participation

Chapter 2 provides information on the incidence of no-shows (children randomly assigned to the Head Start group but who failed to participate) and crossovers (children randomly assigned to the non-Head Start group but who participated in Head Start). These data are provided by age cohort and for both the total sample that was randomly assigned and the subset of children who are part of the Year 1 analysis sample. For this purpose, a child was considered a "no-show" if it was not possible to identify a time when he/she participated in Head Start during the 2002-03 program year after checking several data sources. Similarly, a child was deemed a crossover if information was obtained indicating that he/she participated in Head Start **at any time** during the 2002-03 program year.

The determination of participation status was first based on three sources of information:

- **EnrolOct.** Information reported by site coordinators from a check in early November 2002 with the Head Start centers where random assignment was done on whether a child was enrolled and/or attending the center on or before Oct 15.
- **P3ENROL.** Each parent's response as to whether their child was currently attending Head Start. This information could come from either the fall 2002 (P3ENROL_fall) or the spring 2003 (P3ENROL_spr) parent interview or from both if available.
- **P4EVER.** Each parent's response in fall 2002 as to whether their child had **ever** attended Head Start.

The P3ENROL_fall variable was combined with the P4EVER variable to create a fall parent indicator of Head Start services (Parent_HS_fall). A cross-tabulation was then run on EnrolOct x Parent_HS_fall x P3ENROL_spr, and the results were examined to determine which cases received Head Start over the year and which required further investigation to determine Head Start services (Exhibit A.2.2.1).

Two pieces of corroborating data were needed to make a final determination of each child's Head Start participation. For example, if the Head Start center where random assignment took place reported the child as attending Head Start, the fall 2002 parent interview reported the child in Head Start, but the spring 2003 parent interview reported the child in non-Head Start center-based care, the child was coded as receiving Head Start services at least sometime during the year.

Other cases were more difficult to determine, such as those with a single initial indicator that Head Start participation took place. For example, if the random assignment center reported the child as NOT in Head Start, and if the parent stated that the child received Head Start in fall 2002 and the child was not in Head Start in the spring, additional data were used to determine whether a child actually attended Head Start at any point (see below). There were 322 cases requiring further investigation (i.e., they could not be determined based on the "two corroborating pieces of information" rule.)

For the 322 cases requiring further investigation, the following additional data were examined to determine Head Start participation:

- The type of setting that parents reported their child was currently attending in the fall and/or spring.
- The name of the current setting provided by the parent from the fall and/or spring parent interview.
- Parent-reported dates in fall when the child started and stopped Head Start services.
- Parent-reported dates in fall when the child started and stopped another non-Head Start child care arrangement.
- The location where the child was reported by the study's in-person assessor to be receiving services at the time of the fall and/or spring assessments (N=114).
- The location of the classroom where observation was done for a particular child by the study's in-person assessor in the fall and/or spring (used in only a few cases).

Information obtained from the spring 2003 interview with care providers of children in non-center-based non-parental care at that time.

Exhibit A.2.2.1: Classification of Study Children by Head Start Participation Based on Three Initial Data Sources: All Children Randomly Assigned, Head Start and Non-Head Start Sample Members, and Both Age Cohorts (N=4,667)

Fall 2002 Program Attendance Record	Fall 2002 Parent Interview	Spring 2003 Parent Interview		
		Head Start	**Not Head Start**	**Missing**
Head Start	**Head Start**	Participant (N=1,750)	Participant (N=121)	Participant (N=94)
	Not Head Start	Participant (N=9)	SB (N=31)	SF (N=1)
	Missing	Participant (N=77)	SS (N=7)	Participant (N=78)
Not Head Start	**Head Start**	Participant (N=161)	SB (N=85)	SF (N=23)
	Not Head Start	SB (N=82)	Non-Participant (N=1,205)	Non-Participant (N=94)
	Missing	SS (N=48)	Non-Participant (N=201)	Non-Participant (N=454)
Missing	**Head Start**	Participant (N=37)	SB (N=9)	SF (N=3)
	Not Head Start	SB (N=1)	Non-Participant (N=39)	SF (N=8)
	Missing	SS (N=12)	SS (N=12)	[Impute] (N=25)

<u>Cases requiring further investigation</u>:
SB – checked both fall and spring parent interviews during further investigation (N=208),
SF – checked fall parent interview during further investigation (N=35),
SS – checked spring parent interview during further investigation (N=79).

The process used to investigate these cases was as follows:

- We looked at the fall and/or spring parent interview care setting name, type of setting, Head Start start and end dates (if applicable), and current care provider start dates. In addition, we checked data on the location of fall and/or spring child assessments and from the spring non-parental care provider interview. Based on these data items, we made a determination whether the child received Head Start services (N=294) or concluded that no such determination could be made and imputed participation status instead (N=3).
- We imputed participation status (without checking any other sources) for cases for which all three original data items (i.e., EnrolOct, Parent_HS_fall, and P3ENROL_spr) were missing (N=25).

Appendix 2.3: The Racial/Ethnic Composition of the Study Sample

This appendix examines the distribution of the research sample by race/ethnicity at each stage of sampling for the study, in relation to both the Head Start Program Information Report data system that served as the original starting point for sampling and the newer HSNRS. It shows how the set of newly entering Head Start children studied in this report came to differ somewhat from published information regarding the share of Head Start program participants in different racial/ethnic groups. It also demonstrates that these differences—a somewhat higher share of the age 3 cohort in the Black (non-Hispanic) category than is true for the sampling frame defining the population studied and a somewhat higher share of the age 4 cohort in the Hispanic category—are due largely to normal sampling variation in the selection of the programs, centers, and children for study.

Exhibit A.2.3.1 shows the distribution of the portion of national Head Start enrollment covered by the selected research sample at each stage of the sampling process, by race/ethnicity. It shows a small but steady increase in the percentage of Hispanic children and a small but steady reduction in the percentage of Black children (lines 1 to 10 of the exhibit). There are several causes of the apparent shift in the racial/ethnic distribution:

- Exclusion of programs and Head Start centers in communities saturated by Head Start (i.e., communities where all eligible families interested in Head Start are already served and vacancies exist);
- Chance sampling error when picking the geographically based PSUs at the beginning of the process, as well as in later selection of programs within PSU and centers within program;
- Differences in race/ethnicity reporting procedures among the PIR, HSNRS, and instruments used by the study to measure additional characteristics of individual Head Start centers (the CIF and the applicant rosters); and
- Definitional differences in the populations being compared, newly entering children versus all children served by the program.

Additional deviations occur when the sample is divided by age cohort (lines 11 to 14 of the exhibit). These reflect previously unmeasured variations in the types of children Head Start serves, particularly newly entering children, at different age levels. These various steps are discussed in more detail below.

Racial/Ethnic Distribution at the Program Level (Lines 1 to 6)

The initial sampling frame appears at the top of the exhibit: all grantees in the PIR data system for the 1998-1999 Head Start program year. These were the most recent data available when sampling began in late 2000 (when commitments to include certain sections of the country in the study had to be made to stay on schedule for the research as a whole). Data on race and ethnicity from the PIR are self-reported by agencies and do not break down Head Start enrollees by age reflective of the two analysis cohorts used in this report. As a result, the initial rows of the exhibit provide numbers for the combined group of all children potentially eligible for inclusion in the study.

Line 1 of the exhibit looks at the racial/ethnic composition of the 1,715 Head Start programs (i.e., grantees) that existed in the 1998-1999 program year, with the race/ethnicity data updated to the 1999-2000 program year where feasible,[1] and following PIR data as described by PIR guidelines given to reporting agencies:

- **Actual Enrollment.** "The total number of children who have been enrolled in your program for any length of time, provided they have attended at least one class or, for home-based children, received at least one home visit. This includes children who have dropped out or enrolled late. Those children funded by other sources **who are part of the Head Start program and receive Head Start services** are to be included in the actual enrollment figures."

- **Race/Ethnicity.** "Of the total actual enrollment, the number of children in the following ethnic categories: AMERICAN INDIAN OR ALASKAN NATIVE. (A person having origins in any of the original peoples of North and South America, and who maintains tribal affiliation.); ASIAN. (A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent.); BLACK OR AFRICAN AMERICAN. (A person having origins in any of the Black racial groups of Africa.); HISPANIC OR LATINO. (A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race.); NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER. (A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.); and, WHITE. (A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.)."

Line 2 of the exhibit describes the reduced frame of 355 programs in the 25 PSUs selected for the study (see Chapter 1), weighted by the inverse of each PSU's probability of selection. The source of enrollment data here is again the PIR. The slight increase in the percentage of Hispanic children and decrease in the percentage of Black children at this point are

[1] Updates were not made for programs absent from the 1999-2000 PIR data or with missing data that year.

most likely due to chance sampling errors when selecting the 25 PSUs at random out of a much larger universe of PSUs spanning the entire United States.

Line 3 reduces the frame to just 261 programs through subsampling in the largest PSUs and exclusion of certain unusable programs. These programs are weighted to reflect the multiple steps used to arrive at this frame, i.e., by the product of (1) the inverse of the probability of selection of the PSU, (2) the inverse of the probability of a particular program's being selected when subsampling programs in the largest three PSUs, and (3) an adjustment for excluding from the frame eight programs already involved in extensive data collection for FACES. The source of race/ethnicity data here is again the PIR. Estimates from the subsample of 261 selected programs in line 3 closely match those of the frame from which they were selected (line 2), indicating that chance sampling variation and the eight systematic exclusions did not lead to any shift away from the universe of interest.

Line 4 reflects a frame of 223 programs, dropping from line 3: (1) the programs found to be in saturated communities after screening by study staff and (2) a small number of programs that had closed since the 1998-99 program year. Programs are weighted here as in the previous line: by the product of the inverse of the probability of selection for the PSU, the inverse of the probability of selection for subsampling in the three PSUs, and an adjustment for excluding the eight FACES programs. The source of enrollment data is the PIR. After dropping the saturated and closed programs, the estimates in line 4 remain close to those in line 3 in terms of racial/ethnic composition.

Line 5 of the exhibit looks at the subsample of 90 programs selected for actual participation in the study from among the 223 identified as part of the frame at the previous step. These are weighted by the inverse of each program's overall probability of selection through all steps in the sampling to this point. Using race/ethnicity data from the PIR once again, it is shown that the estimates from the 90 sampled programs very closely match the frame from which they were selected (line 4). Confidence intervals are provided for the estimates of the share of children in each racial/ethnic group, indicating the range of values that almost certainly contains the true overall population share once sampling variation is taken into account. These 95 percent confidence intervals show a fairly wide potential for true population shares to differ from the sample-driven estimates, though the latter remain the single best indicators of how the population represented by the data may have (or in this case, has not) shifted as the set of programs to be studied narrowed.

Line 6 drops three additional saturated programs (ones identified as saturated only once study staff began working with the 90 sampled programs to determine the centers appropriate for inclusion in the study), two that were part of intensive data collection that same year for Head Start's QRC program and one that had closed. Again using race/ethnicity data from the PIR, the percentage of Hispanic children in the universe represented by the data now jumps 3 percentage points, while the percentage of Black children drops by 6 percentage points. The 7 percentage-point difference in the percentage Black enrollment here compared to the first row of the table is probably not due solely to sampling error, since (from the previous step) a 95 percent confidence interval ranges just plus or minus 6 percentage points from its midpoint. Some may be sampling error, however. Still, an analysis of the programs excluded to this point because they are in saturated communities shows them to be less Hispanic than the norm, accounting for another portion of the change. (One excluded saturated program had a very large enrollment that was more than 90 percent Black.) Another contributing factor to the shift in the race/ethnicity distribution may be that the saturation adjustment to the program weights (see Appendix 1.2) does not sufficiently control for race/ethnicity, as race/ethnicity data were only collected for all Head Start enrollees.

Racial/Ethnic Distribution at the Center Level (Lines 7 to 10)

Step 7 of the process moves from programs to centers as the unit of sampling, using the CIF. This form was developed by the study and filled out jointly with program staff to identify and gather information about each center relevant to the sampling process. In total, the 84 remaining programs provided a frame of 1,254 centers. Data on race/ethnicity were collected on the CIF using the same total enrollment concept and the same race/ethnicity categories as the PIR. However, the CIF data in this row are not strictly comparable to PIR information in the previous rows for two reasons: the CIF collected its information 2 to 3 years later than the PIR, and the PIR figures include any child who attended Head Start at any time during the program year involved (1999-2000). The CIF, on the other hand, collected enrollment counts as of a single point in time (October 1, 2001).

Moreover, the CIF data are not 100 percent complete. Eight percent of the centers operated by the 84 sample programs provided no information on their Hispanic enrollment, and 9 percent were missing information on Black enrollment. Values were imputed for these cases by multiplying each center's reported total enrollment by the average percentage Black and percentage Hispanic enrollment for other sampled centers in the same zip code, city, county, program, or PSU (moving outward geographically as far as was required to obtain available data). As a check on the accuracy of the reporting of race/ethnicity enrollments by the centers, the sum of enrollment across all race/ethnicity categories was compared to the reported total enrollment in each center and found to be fairly consistent (i.e., their sum came very close to the reported total enrollment in 95 percent of the centers involved).

Because the line 7 estimates are based on a complete census of all the centers in each sampled program, the sum of enrollments across centers for any program is weighted according to that program's final weight, and no new sampling variability is added. However, the shift in measurement methods results in a slight increase in the percentage Hispanic and a moderate decrease in the percentage Black enrollment. However, these differences appear to be within the overall sampling error of the process to this point, as indicated by the width of the 95 percent confidence intervals (i.e., see line 7).

Line 8 drops saturated centers from the frame but makes no other changes (i.e., estimates are still weighted by the final program weight and the source of enrollment data is the CIF). This produces another slight increase in the percentage of Hispanic enrollees and a further slight decrease in the percentage of Black enrollees as compared with line 7, though enrollee differences are still well within the overall sampling error to this point. However, the total upward "creep" in percentage Hispanic enrollment from the original PIR program frame at step 1 has, by this point, reached 8 percentage points, with an offsetting downward shift in the percentage of Black enrollment of equal magnitude. The percentage of White enrollment is essentially unchanged from its starting point.

Line 9 estimates are based on the final sample of selected centers, 458 of the 1,254 total centers at the previous step, with each center weighted by the inverse of its overall probability of selection (incorporating sampling probabilities at the PSU and program as well as center level). As shown, the race/ethnicity distribution of the sample of centers matches the frame from which it was selected, again based on CIF data.

Line 10 shows the consequences of the removal from the sample of 80 centers because of the late discovery of saturation and closure and, in a very small number of cases, refusal by agency leadership to implement random assignment and participate in the study. Estimates are based on CIF data and weights equal to the inverse of the overall probability of selection of a given center with an adjustment to compensate for the dropped saturated and refusing centers.

This adjustment inflates the weights of the 378 remaining centers to reach the same total enrollment of newly entering children (as reported on the CIF) as the original 458 centers sampled (i.e., the 378 represent all sampled centers still in operation). The net result of the deletions and adjustments is another slight increase in the percentage Hispanic and a slight decrease in the percentage Black enrollment, but again, differences at this particular step are well within the overall sampling error.

Racial/Ethnic Distribution at the Child Level (Lines 11 to 14)

Line 11 shifts again to a new data source, moving progressively closer to counts of enrollees that will actually flow into the frame during child-level sampling (at step 12 below). Once agreement was reached on the exact centers selected for participation in the study and applications for the 2002-2003 program year started to be submitted, grantee and delegate agency staff, supported by the research team, began filling out rosters of all applicants. When assembled on a cumulative basis, these rosters were considered a census of all the Head Start applicants at a given center over the study's intake period and so were weighted using the same final center weights used at the preceding step. The source of the new counts of children by race/ethnicity at this stage was the race/ethnicity field on the pre-formatted roster form, again patterned after the PIR (and hence CIF) categories. Despite using a new data source and a later program year (i.e., the upcoming 2002-2003 year, as compared with the enrollment experiences at the start of the 2001-2002 program year captured by the CIF), the figures for all children on the roster almost identically match those from the earlier CIF data in line 10. The slight shift that does occur continues the very gradual upward creep of the percentage Hispanic and the corresponding incremental decline of the percentage Black enrollments.

The rosters of applicants included information on each child's age group, i.e., whether the child was thought by program staff to be 1 year away from kindergarten entry (the 4-year-old group) or 2 years away (the 3-year-old group). It thus became possible at this stage to conduct sampling separately for the two age groups. Line 11 of the exhibit breaks out the figures for each racial/ethnic group into separate distributions for the two age groups. Eight percent of the children on the rosters had missing data for either age or race/ethnicity and were not included in these figures. As shown in Exhibit A.2.3.1, a major shift occurs when the overall numbers for percentage Hispanic and percentage Black enrollment are broken down into separate numbers by age group: the 3-year-old group is found to be several percentage points more Black and less Hispanic than the average, and the 4-year-old group is several percentage points more Hispanic than the average. These racial/ethnic distinctions by age level have not previously been documented in national Head Start data since the two factors are not cross-tabulated in the PIR.

Another large shift in the populations described also occurs at line 12. Returning children who had already participated in Head Start (or Early Head Start) and a very small number of children considered "high-risk" by participating grantees and delegate agencies were excluded from random assignment based on information on the rosters (a check for duplicate entries further pruned the rosters). Next, using the local agency's eligibility criteria (usually a numerical score), the list of newly entering children that would ordinarily have been enrolled was "extended" to provide for a specified number of children who would subsequently be randomly assigned to the non-Head Start group and not enrolled in the program. (The children added were those who would be "next in line" for admission based on the agency's eligibility criteria.) This extended list became the sampling frame for actual random assignment at step 13 below. Together, the restrictions just described shrank the sampling frame on the rosters from 27,526 children to 14,439 children, with most of the deletions resulting from the exclusion of children who had previously participated in Head Start or Early Head Start.

The line 12 estimates are based on this restricted frame, with each child weighted by the final center weights consistent with the fact that the rosters constituted a census of the relevant children at each center. The source of the race/ethnicity categories is again the rosters of applicants, with the 8 percent of children missing either age or race data excluded from the calculations for Exhibit A.2.3.1 (though not from randomization). Note that as with all applying children, the distribution of race/ethnicity for the "top priority" newly entering applicants in the 3-year-old group differs markedly from that of the 4-year-old group. The percentage Black enrollment is now very much higher for the 3-year-old group than the 4-year-old group, and the percentage Hispanic enrollment follows an equally sharp reverse pattern. For the 3-year-old group, this offsets the gradual drop in percentage Black enrollment in previous rows of the exhibit, while for the 4-year-old group it exacerbates the rise in the Hispanic enrollment. Thus, compared with the race/ethnicity distribution estimated from CIF data just two steps earlier (line 10) the 3-year-old group looks hardly any different (the percentage of Black children has risen 3 percentage points by line 12, mostly at the expense of the percentage of White children). The 4-year-old group distribution has changed radically, however, to 52 percent Hispanic and 17 percent Black enrollment compared to 38 and 28 percent at the earlier point. It is important to recall that the reference population for line 12 is the population of newly entering children, whereas the population for line 10 is the population of all Head Start enrollees.

Line 13 moves from the restricted frame of 14,439 children to the 4,747 children sampled into the Head Start and non-Head Start research groups through random assignment. The randomization algorithm allocated children in the right proportions into statistically equivalent Head Start and non-Head Start samples and into a group of children admitted to Head Start (to provide for full enrollment) but who were not included in the study. The children in the two research samples are banded together in line 13 and weighted by the inverse of each child's overall probability of selection, including child-level sampling at random assignment and all prior stages of selection. The within-center probability of selection was approximated as the ratio of the number of sampled Head Start and non-Head Start children in each center to the newly entering enrollment for the center as a whole. This reflects actual program size rather than the artificial construct of the impact study created by all children included in the random assignment pool. Total newly entering enrollment by center was collected on the CIF as of October 1, 2001, and updated in about half the centers to reflect fall 2002 numbers. The goal was for each research sample to weight up to the national population of newly entering Head Start enrollees as of fall 2002.

The source of the race/ethnicity data for this population is again the application roster. At this point, about 9 percent of child records had missing age or race and were excluded from line 13. The population represented by the selected sample of study children closely matches the frame from which it was selected (line 12)—although the 3-year-old group continues to shift incrementally to greater representation of Black children and less representation of Hispanic children.

The last line in the exhibit, line 14, provides estimates of the population represented by the baseline data used in this report—the sampled Head Start and non-Head Start children for whom cognitive assessments were completed in fall 2002. Each child was weighted by his or her overall probability of selection from the previous step, a nonresponse adjustment to account for children who did not complete fall 2002 assessments, and, for the 4-year-old group, a poststratification adjustment to the race/ethnicity proportions for the newly entering 4-year-olds from the HSNRS. This last adjustment, which could not be done for the 3-year-old group because the HSNRS samples only 4-year-olds, is undertaken to reduce sampling error, as explained below. However, the race/ethnicity data collected by the HSNRS do not follow any type of standardized definitions; they are reported by category using definitions decided by individual grantees and may not be comparable to those of PIR, CIF, or application roster. As before, about 9 percent of the relevant records were missing age or race information on the roster and are excluded from the calculations.

As can be seen by comparing lines 13 and 14, poststratification to the HSNRS race/ethnicity distribution significantly reduces the estimated proportion of Hispanic children and increases the proportion of Black children in the 4-year-old group. This closes most of the gap between the line 13 numbers—the starting point for poststratification—and the control totals for newly entering 4-year-olds shown at the bottom of the exhibit and provided by the HSNRS. Differences remain, however, because the data were not poststratified directly to the overall national distribution for newly entering 4-year-olds from the HSNRS. Rather, for each race/ethnicity category, we poststratified to the ratio of the HSNRS percentage for all programs reporting in the HSNRS to the study sample percentage for 84 programs, using HSNRS first year enrollment data. This poststratification adjustment does not remove real differences in concepts and measurement between the two data sources but **is** intended to reduce the PSU and program component of sampling error (the change from line 1 to line 5 in the Exhibit).

Since the poststratification adjustment closed most of the gap, we learn from this procedure that the difference in the racial/ethnic composition of the 4-year-old group is partially due to sampling error from the sampling of PSUs and programs. Differences in race/ethnicity reporting procedures among the PIR, HSNRS and the NHIS and in the populations being compared (newly entering vs. all children) also contribute to the differences observed. These differences do not necessarily indicate there is a systematic bias in the NHIS sample with respect to race/ethnicity. Presumably the same is true of the 3-year-old group, having been generated in precisely the same manner at every step of the process, described in this appendix.

Exhibit A.2.3.1: Estimates of the Race/Ethnicity Distribution of the Research Sample at Different Stages of Sampling

Data Source - Units Measured	**Observations Examined**	**Percent Hispanic**	**95 Percent Confidence Interval**	**Percent Black**	**95 Percent Confidence Interval**	**Percent White and Other**
1. PIR - Total Enrollment	All programs in the National PIR Data System (N=1,715)	28%		37%		36%
2. PIR - Total Enrollment	Frame of programs in selected PSU's (N=355)	30%		38%		32%
3. PIR - Total Enrollment	Subsample of programs in selected PSU's (N=261)	30%		38%		32%
4. PIR - Total Enrollment	Restricted frame of programs (less saturated) (N=223)	31%		38%		30%
5. PIR - Total Enrollment	Selected programs [grantees/delegate agencies] (N=90)	31%	[25%,36%]	39%	[33%,45%]	30%
6. PIR - Total Enrollment	Final sample of programs (N=84)	34%		33%		33%
7. CIF - Total Enrollment	Frame of centers (N=1,423)	35%	[30%,40%]	30%	[24%,36%]	34%
8. CIF - Total Enrollment	Restricted frame of centers (less saturated) (N=1,254)	36%	[31%,41%]	29%	[24%,34%]	35%
9. CIF - Total Enrollment	Selected centers (N=458)	36%	[30%,42%]	29%	[23%,35%]	35%
10. CIF - Total Enrollment	Selected centers where RA was done (N=378)	38%	[32%,44%]	28%	[22%,34%]	34%
11. Roster - All Applicants	Frame of children (including exempt) (N=27,562)	39%		27%		34%
	All 3-Year-Old Group	*35%*		*31%*		*34%*
	All 4-Year-Old Group	*42%*		*25%*		*33%*
12. Roster:	Restricted frame of children (N=14,439)	***44%***		***24%***		***32%***
Nonexempt, Newly entering Applicants	Newly Entering 3-Year-Old Group	***37%***		***31%***		***32%***
	Newly Entering 4-Year-Old Group	***52%***		***17%***		***31%***
13. Roster:	Sampled children (N=4,747)	***43%***		***25%***		***32%***
Nonexempt, Newly entering Applicants	Newly Entering 3-Year-Old Group	***34%***		***33%***		***33%***
	Newly Entering 4-Year-Old Group	***53%***		***16%***		***31%***
14. Roster:	Poststratification (N=3,723 child respondents)	***37%***		***30%***		***33%***
Nonexempt, Newly entering Applicants	Fall 2002 NHIS - Newly Entering 3-Year-Old Group	***33%***		***35%***		***32%***
	Fall 2002 NHIS - Newly Entering 4-Year-Old Group w/ poststratified weight	***43%***		***24%***		***33%***
	HSNRS: Newly Entering 4-Year-Old Group	36%		25%		39%
	HSNRS: Returning 4-Year-Old Group	26%		35%		39%

Appendix 3.1: Differences Between Main Arrangement and Focal Arrangement

Definition of Main Arrangement
The arrangement in which the child spends the most time between the hours of 9 a.m. and 3 p.m. Monday through Friday. Head Start is always defined as the main arrangement for children enrolled in Head Start.

When compared to the focal arrangements in Exhibit 3.2 in Chapter 3, the main arrangements presented in this appendix (see Exhibit A.3.1 below) indicate that the differences in the proportion of children with particular focal and main arrangements are relatively small. Furthermore, because Head Start is defined as both the main and focal arrangement—independent of hours per week—for children enrolled in Head Start, the proportions of children with Head Start as their main and focal arrangements are identical.

Exhibit A.3.1: Percentage of Children in Head Start and Control Groups by Type of Main Care Arrangement in Spring 2003, Weighted Data

Type of Arrangement	Head Start Treatment Group (sample size=1,336)	Control Group (sample size=821)	Head Start Treatment Group (sample size=1,068)	Control Group (sample size=662)
	Percent of 3-year-old group		Percent of 4-year-old group	
Parent Care	**7.8*****	**42.6**	**10.7*****	**48.9**
Non-Parental Care	**92.2*****	**57.4**	**89.3*****	**51.1**
Head Start	84.1***	17.5	76.4***	13.4
Non-Head Start Center	6.6***	23.7	11.1***	27.6
Non-Relative's Home	0.3***	5.6	0.7**	4.7
Relative's Home	0.7***	7.1	0.5*	3.2
Child's Home w/Relative	0.5**	3.3	0.5*	2.1
Child's Home w/Non-relative	0.0	0.2	0.2	0.1
Total percent	**100%**	**100%**	**100%**	**100%**

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

The small differences between focal and main arrangements among children not enrolled in Head Start arise from two types of circumstances. First, some children who were mainly in parent care also spent at least 5 hours per week in a non-parental preschool or child care arrangement. When considering the arrangement in which children spent the **most** time (rather than the focal arrangement) the percentage of children in parent care only increases by 1 to 2 percentage points among children assigned to the Head Start group and 4 to 7 percentage points among children in the non-Head Start group. Second, some children, mainly in small, home-based, non-parental preschool and child care arrangements (such as care by a relative in their own home), also spent at least 5 hours per week in an arrangement that we hypothesized might be more likely to offer the types of educational, social, and access-to-services opportunities offered by Head Start. Again, the proportion of children in this situation is relatively small, so the differences between main and focal arrangements are not particularly substantial.

We explored two different definitions of children's weekday arrangements, main and focal, to better understand the patterns of preschool and child care use among families in our sample. This allows future analyses to consider various definitions of the counterfactual, i.e., the alternatives to Head Start used by families in our sample. In general, however, this report relies on focal arrangements in describing the treatment for children assigned to the Head Start group and the alternative to the treatment for families assigned to the non-Head Start group.

Appendix 4.1: Imputations for Item Nonresponse in the Fall 2002 Data

To facilitate analysis of the data, and to ensure that the results obtained by different analysts are consistent with one another, it is desirable to impute missing responses to produce as complete a data set as possible. Imputation also helps to control for nonresponse bias and produce a more representative file for analysis. For example, many software packages select only the cases that are complete on the set of variables analyzed and ignore the cases with incomplete data. Discarding incomplete cases is inefficient, but more seriously, the complete cases may not be representative of the target population; consequently, estimates derived from them are subject to nonresponse bias.

For this study, missing values for fall 2002 variables due to item nonresponse were imputed using hot deck imputation. Hot deck imputation is a procedure where cases with missing values for specific variables have the “holes” in their records filled in with values from other similar cases. Because the imputed values come from actual respondents’ values, hot-deck imputation has the desirable property that imputed values are always realistic and preserve the underlying sampling variation in the data.

The “donor” case from which the imputed value is taken (also referred to as the respondent), is randomly selected from a pool of similar children who are matched to the “recipient” (or nonrespondent) on characteristics that are correlated with the variable being imputed. The aim is to construct pools (or imputation classes) that explain as much of the variance in the variable to be imputed as possible, but are of adequate size so that there is some minimum number of respondents in each class, and donors are not reused too many times. The assumption is that within each imputation class, the mechanism that leads to missing data is “ignorable”; that is, the missing values are as though they were missing at random. This means that the probability that a value is missing can depend on the values of the imputation class variables but, within class, not on the missing outcome values. If implemented carefully, hot deck imputation can preserve the distribution of the data on measured variables so that estimates of distributional characteristics such as percentiles, variability, and correlation will not be distorted. However, if the item response rate is very high, a small percentage of imputed data will have very little effect on the distribution of the variable regardless of the imputation method.

The variables used to form imputation classes or cells were identified from chi-square tests of association and bivariate correlation coefficients. In some cases, they were also determined by skip patterns in the parent questionnaire and other requirements of logical consistency between questionnaire items. The imputation cells were created by cross-tabulating all of these variables at once. A donor was allowed to be used up to three times. When no more donors were available in an imputation class, adjacent cells were collapsed. The order of collapsing was specified so that levels of the least correlated cell variable were collapsed first, followed by the second least correlated variable, etc. until a donor was found. Imputed values have been flagged so that an analyst has the option of not using the imputed data, such as when analyzing the effects of the imputed data on the results.

We imputed missing data for all fall 2002 demographic variables and the fall 2002 measures of each of the spring 2003 outcomes (e.g., parenting practices, child health, assessment scores, child socio-emotional behavior, and other scale variables). The variables that underwent imputation and their item nonresponse rates for the analysis sample used in this report (the spring 2003 child assessment respondents) are given in Exhibit A.4.1.1.

The logical relationships between items were taken into account in the imputation to maintain consistency of the data and attempt to preserve correlations among variables. Closely correlated items such as assessment scores or socio-emotional scales were usually imputed from a single donor child. The donor was randomly selected from within a donor pool of children matched by treatment/control group assignment, language spoken at home, sex, race/ethnicity, and age in months as of September 1, 2002. The score and scale variables were imputed in groups according to similar patterns of missingness (i.e., the joint missing rates) and the degree of correlation among them. This strategy was viewed as a compromise between the desire to avoid throwing away reported scores and the goal of preserving the correlation among score variables. In general only the missing scores were imputed on each record, and children with partially reported scores did not have them overwritten by the donor's scores. However, for patterns of missingness represented by a small number of children, the donor's scores were allowed to overwrite the reported scores in the interests of reducing the number of computer runs. It should be noted that the percentage of child records with partial reporting of score and scale variables is small. The socio-emotional scales were either entirely missing or entirely reported for all but a trivial (< 0.1%) percentage of the sample. For the depression, locus of control, welfare, and crime and violence scales, 8.3 percent of the sample had partially missing data (5.6 percent were missing all but one scale, 2.5 percent were missing only one scale, and 0.2 percent were missing some other combination). For the continuous score variables, less than 5 percent of the sample had partial reporting of scores; most were either missing all scores or none.

The order in which items are imputed is also important in preserving the correlation structure in the data, because some imputed items can be used to form imputation cells in the subsequent imputation of related items. This strategy was used, for example, in the imputation of categorical assessment scores, so that the first score that was imputed could be used to create imputation cells for the next score. It was also used throughout in the imputation of correlated demographic and household variables. Similarly, for items associated with a skip pattern in the parent questionnaire, the item that leads into the skip pattern was imputed first and the subsequent items were imputed depending on the value of the skip indicator. The demographic variables were imputed first, then used to impute parenting practice, household income, child health, assessment score, and scale variables. Items with the least amount of nonresponse within a group of related categorical variables were imputed first, then used in the imputation of items with larger amounts of missing data.

In general, donors were randomly selected from within the same Head Start program within a cell when possible, collapsed with a geographically adjacent program in the cell when necessary. Programs were sorted within a cell by broad geographic area (our primary sampling unit, or PSU) within Census region, so adjacent programs tended to be from the same county or a nearby county. When there were a large number of imputation cells, the donor search often was broadened to the entire geographic PSU within a cell, and sometimes PSUs within a region were also collapsed. Some items such as fall scores required a closer match on demographic variables than geography or Head Start program in order to find a similar donor pool, and no attempt to stay within the PSU or program was made for these. Geography was also ignored for certain items requiring a very close match to the donor on other questionnaire items for logical consistency.

The distribution of each imputed variable was compared before and after imputation to check that the imputation procedures had not appreciably changed the distribution of the variable. Correlation matrices were examined to check that bivariate correlations among scores and scales were not attenuated. Crosstabs between categorical variables involved in skip patterns and those requiring logical consistency were checked to make sure that inconsistencies had not been introduced. The only variable where the distribution shifted more than a trivial amount was father's employment status, which had a very high missing rate of 51 percent. The percent age of fathers employed full-time shifted from 74 percent to 71 percent, and the percentage unemployed increased from 16 percent to 20 percent. Fathers for whom employment status is unknown tend to come from cells with higher unemployment rates among respondents; thus, the inclusion of their imputed values will raise the overall unemployment rate. The variables used to create imputation classes for employment status were receipt of food stamps, receipt of TANF, father's level of education, father's race, and PSU.

Exhibit A.4.1.1: Item Nonresponse Rates for Imputed Variables

Variable Name	Reported Count	Imputed Count	Percent Imputed	Total of Reported and Imputed Count
Crime & Violence Maximum Likelihood Ability Estimate	3,546	352	9.0%	3,898
Crime & Violence IRT True-Score	3,546	352	9.0%	3,898
Number of children age 17 and under in household	3,796	102	2.6%	3,898
Restricting Child Movement Scale – fall	3,539	359	9.2%	3,898
Family Cultural Enrichment Scale	3,524	374	9.6%	3,898
Family Cultural Enrichment Scale 2	3,540	358	9.2%	3,898
Removing Harmful Objects Subscale – fall	3,538	360	9.2%	3,898
# Times child is read to	3,548	350	9.0%	3,898
Safety Devices Subscale – fall	3,538	360	9.2%	3,898
Parental Safety Practices Scale – fall	3,537	361	9.3%	3,898
Spanked child in last week	3,544	354	9.1%	3,898
# Times spanked child	3,528	370	9.5%	3,898
Used time out in last week	3,542	356	9.1%	3,898
# Times used time out	3,524	374	9.6%	3,898
Adult books in home	3,547	351	9.0%	3,898
Derived caregiver's race	141	7	4.7%	148
Derived child race	3,882	16	0.4%	3,898
Child sex	3,898	0	0.0%	3,898
Derived father's race	3,710	188	4.8%	3,898
Head Start participation	3,897	1	0.0%	3,898
Derived mother’s race	3,777	121	3.1%	3,898
Caregiver’s age	137	11	7.4%	148
Child born in the United States	3,792	106	2.7%	3,898
Economic difficulty scale	3,525	373	9.6%	3,898
Father’s employment status	1,875	2023	51.9%	3,898
Father’s highest educational attainment	3,460	438	11.2%	3,898
Father’s marital status	3,421	477	12.2%	3,898
Father’s age	3,283	615	15.8%	3,898
Biological father’s immigrant status	3,702	196	5.0%	3,898
Biological father a recent immigrant	1,273	90	6.6%	1,363
Biological father lives with child	3,660	238	6.1%	3,898
Biological father years in the United States	1,170	193	14.2%	1,363
Grandparent in the household	3,786	112	2.9%	3,898
Anyone in household with health condition	3,537	361	9.3%	3,898
Homelessness	3,535	363	9.3%	3,898
Primary home language	3,870	28	0.7%	3,898
Biological mother’s immigrant status	3,773	125	3.2%	3,898
Biological mother recent immigrant	1,210	104	7.9%	1,314
Biological mother lives with child	3,789	109	2.8%	3,898
Biological mother years in the United States	1,210	104	7.9%	1,314
Household monthly income range	3,403	495	12.7%	3,898
Mother’s employment status	3,598	300	7.7%	3,898

Exhibit A.4.1.1: Item Nonresponse Rates for Imputed Variables (continued)

Variable Name	Reported Count	Imputed Count	Percent Imputed	Total of Reported and Imputed Count
Mother has a GED	3,757	141	3.6%	3,898
Biological mother educational attainment	3,757	141	3.6%	3,898
Mother's marital status	3,759	139	3.6%	3,898
Mother's age	3,722	176	4.5%	3,898
Number of moves in last 12 months	3,449	449	11.5%	3,898
Other caregiver's employment status	135	13	8.8%	148
Other caregiver's educational attainment	134	14	9.5%	148
Number of adults 18 and over in household	3,534	364	9.3%	3,898
Primary caregiver health impairs caring for child	3,545	353	9.1%	3,898
Primary caregivers health	3,545	353	9.1%	3,898
Child had dental care, fall 02	3,542	356	9.1%	3,898
Child's health status, fall 02	3,544	354	9.1%	3,898
Child had care for an injury, fall 02	3,537	361	9.3%	3,898
Child has health insurance, fall 02	3,542	356	9.1%	3,898
Child needs ongoing health care, fall 02	3,785	113	2.9%	3,898
Child has regular place for medical care, fall 02	3,538	360	9.2%	3,898
PELS, fall 02	3,548	350	9.0%	3,898
Child has special needs, fall 02	3,787	111	2.8%	3,898
Child has an unmet health need, fall 02	3,540	358	9.2%	3,898
Housing problems scale	3,514	384	9.9%	3,898
Receives Food Stamps	3,771	127	3.3%	3,898
Receives TANF	3,765	133	3.4%	3,898
Respondent's relationship to child	3,786	112	2.9%	3,898
Public or subsidized housing	3,523	368	9.5%	3,891
Mother had a teen birth	3,733	165	4.2%	3,898
Number of children under age 6 in household	3,796	102	2.6%	3,898
Depression maximum likelihood ability estimate	3,536	362	9.3%	3,898
Depression IRT true-score	3,536	362	9.3%	3,898
Elision IRT score	2,408	294	10.9%	2,702
Elision true score	2,408	294	10.9%	2,702
PPVT IRT score	3,187	465	12.7%	3,652
PPVT true score	3,187	465	12.7%	3,652
PPVT standard score	3,187	465	12.7%	3,652
PPVT W-ability score	3,187	465	12.7%	3,652
Spanish Elision IRT score	1,015	124	10.9%	1,139
Spanish Elision true score	1,015	124	10.9%	1,139
TVIP IRT score	1,038	101	8.9%	1,139
TVIP true score	1,038	101	8.9%	1,139
TVIP standard score	1,038	101	8.9%	1,139
TVIP W-ability score	1,038	101	8.9%	1,139
Locus of control IRT scale score	3,534	364	9.3%	3,898
Locus of control true scale score	3,534	364	9.3%	3,898
Is respondent mother or father?	3,796	102	2.6%	3,898
How well did child do in bear counting	3,434	464	11.9%	3,898

Exhibit A.4.1.1: Item Nonresponse Rates for Imputed Variables (continued)

Variable Name	Reported Count	Imputed Count	Percent Imputed	Total of Reported and Imputed Count
Bear counting score	3,260	638	16.4%	3,898
Book score, total	3,473	425	10.9%	3,898
Color name score, total	3,516	382	9.8%	3,898
CTOPPP Elision total score	2,408	294	10.9%	2,702
CTOPPP Spanish Elision total score	1,015	124	10.9%	1,139
CTOPPP print score	1,034	105	9.2%	1,139
McCarthy total drawing score	3,508	390	10.0%	3,898
KFAST raw score	3,220	678	17.4%	3,898
PPVT: total score	3,187	465	12.7%	3,652
Print knowledge score: total	3,445	453	11.6%	3,898
TVIP: total score	1,038	101	8.9%	1,139
WJ3 Applied problems standard score	2,392	310	11.5%	2,702
S_WJ3APPLIED_W	2,392	310	11.5%	2,702
WJ3 Applied problems W score	2,378	324	12.0%	2,702
WJ3 Oral comprehension standard score	2,378	324	12.0%	2,702
WJ3 Oral comprehension W score	2,426	276	10.2%	2,702
WJ3 Spelling W score	2,426	276	10.2%	2,702
WJ3 Letter-word standard score	3,217	435	11.9%	3,652
WJ3 Letter-word W score	3,217	435	11.9%	3,652
WJ3 Applied problems total score	2,392	310	11.5%	2,702
WJ3 Oral comprehension, total score	2,378	324	12.0%	2,702
WJ3 Spelling, total score	2,426	276	10.2%	2,702
WJ3 Letter-word total score	3,217	435	11.9%	3,652
WM Applied problems total score	1,017	122	10.7%	1,139
WM Applied problems, standard score	1,017	122	10.7%	1,139
WM Applied problems, W score	1,017	122	10.7%	1,139
WM Dictation, total score	1,024	115	10.1%	1,139
WM Dictation, standard score	1,024	115	10.1%	1,139
WM Dictation, W score	1,024	115	10.1%	1,139
WM Letter-word, total score	1,028	111	9.7%	1,139
WM Letter-word, standard score	1,028	111	9.7%	1,139
WM Letter-word, W score	1,028	111	9.7%	1,139
Child age as of 9/1/02	3,898	0	0.0%	3,898
Welfare IRT scale score	3,689	209	5.4%	3,898
Welfare true scale score	3,689	209	5.4%	3,898

Appendix 4.2: Comparison of Weighted and Unweighted Mean Differences by Age Cohort

Table A.4.2.1: Spring Outcomes: Treatment and Control; Weighted and Unweighted Cohort 3

OUTCOME	Weighted			Unweighted		
	Treatment	**Control**	**Difference**	**Treatment**	**Control**	**Difference**
Cognitive Domain						
Peabody Picture Vocabulary Test (PPVT)	254.01	249.98	4.02	254.09	250.53	3.56
CTOPPP Elision	243.37	239.73	3.64	244.52	242.25	2.27
Letter Naming Task	5.49	3.92	1.57	5.55	4.14	1.41
Color Naming	13.94	13.03	0.91	13.87	12.80	1.07
Counting Bears	2.85	2.68	0.16	2.86	2.65	0.21
McCarthy Scales of Children's Abilities Draw-a-Design Subtest	3.23	3.04	0.18	3.25	3.08	0.18
Woodcock-Johnson III Letter-Word Identification	306.98	300.51	6.46	307.13	301.16	5.97
Woodcock-Johnson III Spelling	346.57	343.64	2.93	347.52	344.21	3.31
Woodcock-Johnson III Applied Problems	377.25	373.57	3.68	376.91	373.62	3.30
Woodcock-Johnson III Oral Comprehension	435.52	435.44	0.08	435.25	435.54	-0.29
Test de Vocabulario en Imagenes Peabody (TVIP)	254.22	246.00	8.22	257.26	252.04	5.23
Batería Woodcock-Muñoz Revisada: Identificación de letras y palabras	352.13	350.20	1.93	351.91	350.68	1.23
Parent (reported) Emergent Literacy Scale (PELS)	2.87	2.36	0.51	2.94	2.40	0.54
Social-Emotional Domain						
Social Skills and Positive Approaches to Learning	12.42	12.37	0.05	12.43	12.39	0.03
Total Child Behavior Problems Scale	5.77	6.26	-0.48	5.79	6.27	-0.48
Aggressive Behavior Scale	2.96	3.05	-0.09	2.94	3.08	-0.13
Hyperactive Behavior Scale	1.69	2.01	-0.32	1.72	2.04	-0.32
Withdrawn Behavior Scale	0.55	0.58	-0.03	0.58	0.56	0.02
Social Competencies Checklist	10.96	10.99	-0.03	10.96	10.96	-0.00

Table A.4.2.1: Spring Outcomes: Treatment and Control; Weighted and Unweighted Cohort 3 (continued)

OUTCOME	Weighted			Unweighted		
	Treatment	Control	Difference	Treatment	Control	Difference
Parenting Practices Domain						
Parent used time out in the last week	0.62	0.66	-0.04	0.62	0.66	-0.04
Number of times parent used time out in the last week	1.64	1.91	-0.27	1.66	1.89	-0.23
Parent spanked child in the last week	0.41	0.48	-0.07	0.41	0.47	-0.06
Number of times parent spanked child in the last week	0.81	0.99	-0.17	0.82	0.99	-0.17
Parental Safety Practices Scale	3.73	3.69	0.03	3.72	3.71	0.01
Removing Harmful Objects Scale	3.91	3.87	0.03	3.90	3.89	0.01
Restricting Child Movement Scale	3.88	3.89	-0.01	3.88	3.91	-0.03
Safety Devices Scale	3.38	3.30	0.08	3.37	3.33	0.04
Family Cultural Enrichment Scale	3.77	3.54	0.23	3.81	3.58	0.23
How many times child was read to in the last week by parent or other family member	2.94	2.77	0.17	2.93	2.78	0.15
Health Domain						
Child seen by dentist since last September	0.69	0.52	0.17	0.66	0.51	0.14
Overall child’s health status	0.81	0.76	0.05	0.82	0.78	0.03
Child has injury in last month requiring medical treatment	0.09	0.08	0.01	0.10	0.11	-0.01
Child has health insurance	0.92	0.92	0.00	0.92	0.91	0.02
Child has place for routine medical care	0.98	0.98	0.01	0.98	0.98	0.00
Child has condition that requires ongoing medical care	0.13	0.13	0.00	0.14	0.13	0.01
Child has an unmet health care need	0.01	0.02	-0.01	0.01	0.02	-0.01

Table A. 4.2.2: Spring Outcomes: Treatment and Control, Weighted and Unweighted Cohort 4

OUTCOME	Weighted			Unweighted		
	Treatment	Control	Difference	Treatment	Control	Difference
Cognitive Domain						
Peabody Picture Vocabulary Test (PPVT)	293.87	291.34	2.53	292.89	290.29	2.61
CTOPPP Elision	275.24	273.65	1.59	274.38	272.90	1.48
Letter Naming Task	11.53	9.21	2.33	11.10	8.98	2.12
Color Naming	17.10	16.45	0.65	16.88	15.94	0.94
Counting Bears	3.76	3.60	0.16	3.74	3.55	0.19
McCarthy Scales of Children's Abilities Draw-a-Design Subtest	4.55	4.38	0.17	4.55	4.34	0.21
Woodcock-Johnson III Letter-Word Identification	325.46	319.22	6.24	324.44	318.66	5.78
Woodcock-Johnson III Spelling	371.56	367.67	3.89	371.35	368.15	3.20
Woodcock-Johnson III Applied Problems	397.47	394.42	3.05	396.75	393.58	3.17
Woodcock-Johnson III Oral Comprehension	443.40	443.65	-0.24	442.82	443.42	-0.60
Test de Vocabulario en Imágenes Peabody (TVIP)	295.61	291.59	4.02	302.28	293.27	9.01
Batería Woodcock-Muñoz Revisada: Identificación de letras y palabras	358.01	357.08	0.92	359.39	357.72	1.67
Parent (reported) Emergent Literacy Scale (PELS)	3.77	3.33	0.44	3.76	3.31	0.45
Social-Emotional Domain						
Social Skills and Positive Approaches to Learning	12.46	12.49	-0.03	12.51	12.55	-0.04
Total Child Behavior Problems Scale	5.58	5.83	-0.26	5.79	5.88	-0.09
Aggressive Behavior Scale	2.72	2.87	-0.15	2.80	2.86	-0.06
Hyperactive Behavior Scale	1.69	1.79	-0.09	1.78	1.85	-0.07
Withdrawn Behavior Scale	0.66	0.70	-0.04	0.69	0.69	0.00
Social Competencies Checklist	11.03	11.06	-0.03	11.07	11.07	0.00

Table A. 4.2.2: Spring Outcomes: Treatment and Control, Weighted and Unweighted Cohort 4 (continued)

OUTCOME	Weighted			Unweighted		
	Treatment	Control	Difference	Treatment	Control	Difference
Parenting Practices Domain						
Parent used time out in the last week	0.61	0.68	-0.08	0.62	0.66	-0.04
Number of times parent used time out in the last week	1.71	1.66	0.06	1.81	1.57	0.24
Parent spanked child in the last week	0.37	0.37	0.00	0.37	0.39	-0.01
Number of times parent spanked child in the last week	0.71	0.68	0.03	0.72	0.71	0.00
Parental Safety Practices Scale	3.73	3.71	0.03	3.74	3.70	0.04
Removing Harmful Objects Scale	3.89	3.88	0.00	3.89	3.88	0.01
Restricting Child Movement Scale	3.88	3.87	0.01	3.89	3.86	0.03
Safety Devices Scale	3.42	3.36	0.06	3.43	3.35	0.08
Family Cultural Enrichment Scale	4.00	3.91	0.10	3.94	3.85	0.09
How many times child was read to in the last week by parent or other family member	2.96	2.80	0.16	2.94	2.83	0.12
Health Domain						
Child seen by dentist since last September	0.73	0.57	0.16	0.72	0.55	0.17
Overall child's health status	0.79	0.81	-0.02	0.80	0.81	-0.01
Child has injury in last month requiring medical treatment	0.12	0.12	-0.00	0.12	0.11	0.01
Child has health insurance	0.89	0.88	0.01	0.88	0.87	0.00
Child has place for routine medical care	0.97	0.97	-0.00	0.98	0.97	0.01
Child has condition that requires ongoing medical care	0.11	0.11	0.00	0.12	0.11	0.00
Child has an unmet health care need	0.03	0.04	-0.01	0.03	0.03	0.00

Appendix 4.3: Impact Regression Procedures

This appendix describes the regression procedures used to calculate and test the statistical significance of all impact estimates in the report, building on the less detailed discussion of this topic in Chapter 4. It begins by explaining the analysis samples used before proceeding through the different impact regression models from the simplest to the most complex: regressions without covariates, regressions with covariates, and regressions with interactions to examine subgroup and moderator effects.

Analysis Samples Used

The unit of analysis for all impact analyses is the child. This is true irrespective of the outcome measure or data source considered; even outcomes reported by parents and caregivers (the majority) are weighted and analyzed according to the children they described. This makes all impact findings representative of all Head Start children in the nation in 2002. The child weights applied during analysis (see Appendix 1.2) make each child in this universe count equally, not each parent/caregiver nor each Head Start center nor each grantee/delegate agency. Weighting adjustments are made to account for the exclusion from the frame of "saturated" programs and centers.

Different collections of observations are used for different impact purposes. The most important variation concerns the division of the available spring 2003 data into two distinct age-level cohorts: all findings are derived and presented separately for the age 3 cohort and the age 4 cohort. A given cohort is split further into language groups for some purposes, based on the primary language used in the initial cognitive assessments of children in fall 2002 (Spanish versus English + Other). Very rarely—for certain subgroup and moderator analyses noted elsewhere in the report—a sample is deliberately restricted below this level such as when any child with a deceased biological parent is excluded from the analysis of parents' marital status as a moderator. Further small variations occur due to missing data in the spring 2003 outcome observation period, described for cognitive outcomes derived from in-person child assessments and all other outcomes taken from interviews with parents and primary caregivers (hereafter referred to for brevity simply as "parents").

For most cognitive outcomes, impact estimates are calculated using data from all children assessed in spring 2003.[1] This gives a nearly identical sample for all findings on a given age cohort, and where relevant, language group, with variations based on the inability to compute all desired test scores for all assessed children; a small number of cases within the common sample of all assessed children had to be omitted on an outcome-by-outcome basis for this reason. Similar slight variations within a uniform basic sample occur for outcomes measured in parent interviews. Here, item nonresponse in otherwise completed interviews creates case-by-case omissions from an otherwise uniform sample in a small number of cases.

Variations in the analysis sample across data collection instruments are more common due to assessed children without parent interviews and parent interviews without child assessments. Total sample sizes (i.e., number of respondents) for each data type are provided in Exhibit A.4.3.1, as is information on the extent of overlap between the parent interview sample and the child assessment sample. Overlap is considerable for both age cohorts, all language groups, and sample sizes, prior to the outcome-by-outcome exclusions described above, and track closely between the two different data sources. There are only two ways to move closer to a single, totally uniform sample for each age cohort so that impacts on all outcomes would derive from exactly the same set of cases: impute missing outcomes (and entirely missing data collection instruments) for cases with available data for some but not all outcome measures or choose not to use data that are available by excluding observations with less than universal spring 2003 data from all analyses. We do neither of these: the latter would waste information while cutting sample sizes unnecessarily (if still only modestly), while the former would require assumptions too closely intertwined with the program impacts the study intends to measure with observed data not imputed values.

Exhibit A.4.3.1: Number of Respondents in the Analysis Sample, by Data Collection Instrument

	Parent Interview		
Child Assessments	**Respondents**	**Nonrespondents**	**Total**
Respondents	3,808	90	3,898
Nonrespondents	79	690	769
Total	3,887	780	4,667

[1] The exception is the PELS, which comes from parent interviews and fits under discussion of that instrument and its sample definitions.

Regressions Without Covariates

For continuous outcome variables (e.g., PPVT III adapted scale score), impact estimates are based on ordinary least-squares (OLS) regression models applied to the weighted data[2] that replicate the difference-in-means calculation by expressing spring 2003 outcomes as the sum of an intercept term and a shift in the intercept produced by a dummy variable for inclusion in the Head Start group. Using *Y* to represent the outcome measure, this equation is

$$Y = A + BH ,$$

where *H* is a dummy variable for inclusion in the Head Start group. The estimated coefficients from this model, *a* and *b* (estimating *A* and *B* respectively) have the following equivalence to calculated measures from the difference-in-mean approach:

a = *Y bar (c)*

b = *Y bar (h)* - *Y bar (c)* ,

where *Y bar (h)* is the weighted mean value of *Y* for the Head Start sample and *Y bar (c)* is the weighted mean value of *Y* for the non-Head Start comparison sample. By either formulation, *b* gives an estimate of the impact of access to Head Start unbiased by selection into and out of the program, since no systematic differences can exist between the two samples (assuming complete follow-up data on *Y*) given both were chosen as random subsamples of all children randomly assigned and hence the universe of interest (the national population of newly entering Head Start children in communities with more potential Head Start participants than funded Federal Head Start slots).

When divided by its standard error, *b* follows the students' t distribution with 51 degrees of freedom under the null hypothesis that true impact, *B*, is 0, where 51 is the number of degrees of freedom associated with the jackknife estimate of the variance of *b*. (This makes the usual OLS assumption that the dependent variable, *Y*, and hence all estimated coefficients, have normal distributions.) An unbiased standard error for *b* reflective of how the sample was drawn and weighted is obtained using replicates and weights described in Appendix 1.2. The last step conducts a two-tailed test of the null hypothesis of no Head Start impact (i.e., *B=0*) to allow the possibility of program effects in either direction, up or down. Three different levels of statistical

---

[2] See Appendix 1.2 for a description of how analysis weights and replicate weights were constructed based on initial probabilities of selection at different levels of sampling, adjustments for follow-up data nonresponse, and raking to external control totals.

significance, i.e., three different probabilities of rejecting a true null hypothesis, are used and reported in the tables of results in the body of the report, 0.05, 0.01, and 0.001.[3]

Logistic regressions are used in place of OLS regressions for discrete (0/1) outcome variables such as the use of dental care. Here, the specification is non-linear to accommodate the non-normal distribution of Y, which must always take on a value of 0 or 1. However, the model can essentially parallel that of continuous variables with a non-linear transformation added. Specifically, the model expresses the natural log of the odds ratio—the probability that $Y=1$ divided by the probability $Y=0$—as the sum of an intercept term and a shift in the intercept produced by a dummy variable for inclusion in the Head Start group:

$$ln\,[P/(1-P)] \;=\; C + DH,$$

where P is the probability that $Y = 1$ and, hence, $1-P$ is the probability $Y = 0$. The coefficients in this model, C and D, are estimated using a maximum-likelihood statistical routine in the SUDAAN software package that again takes appropriate account of the complex sampling and weighting structure of the data. Standard errors for these estimates are derived using jackknife replication.

Though it occupies a position similar to B's in this model, the coefficient D does not replicate the difference-in-means calculation on Y. It does, however, capture the difference in the typical outcome between the Head Start sample and the non-Head Start sample in an appropriate fashion, calibrated in the non-natural units of log-odds. Once C and D are estimated as c and d, respectively, the meaning of d can be recovered in more intuitive units that show it to be the impact of access to Head Start on the probability of a positive outcome, such as a dental visit occurring. Favorable results in this respect translate into more frequent occurrence of the desired outcome in the real world—a positive Head Start impact. To make the translation out of log-odds space, we calculate the difference between two quantities:

- The log-odds ratio for children in the Head Start group, $c + dH,$ converted into the probability of a positive outcome given access to Head Start by the inverse transformation,

  $P(H=1) = exp\,(c+d) / [1 + exp\,(c+d)]$, and

[3] Operationally, the set of tests is accomplished by determining whether the calculated probability of obtaining the observed impact estimate b when $B=0$—known as the "p value" of the estimate—falls below the 0.05, 0.01, and/or 0.001 significance levels of the respective tests.

- the log-odds ratio for children in the non-Head Start group, c, converted into the probability of a positive outcome absent access to Head Start by the same transformation,

$$P(H=0) = \exp(c) / [1 + \exp(c)] .$$

Hence, the impact estimate from a logistic model is

$$P(H=1) - P(H=0) = \exp(c+d) / [1 + \exp(c+d)] - \exp(c) / [1 + \exp(c)] .$$

Logically, this quantity differs from 0 if and only if d differs from 0. Thus, if assignment to the Head Start group has a statistically significant impact on the log-odds ratio through D (as estimated by d) when tested using the maximum-likelihood assumptions of the logistic model, we can conclude that it also significantly influences the probability of a favorable outcome, $P(H=1) - P(H=0)$. Significance test results reported in the tables in the body of the report are determined in this manner.

Adding Controls for Fall 2002 Factors

The linear and logistic impact models described above can be extended to include fall 2002 characteristics of children and families as predictors of spring 2003 outcomes. The addition of these covariates is represented through the addition of a set of background variables, represented here collectively by the symbol X, to the models already presented:

$$Y = A + BH + EX$$

for continuous outcome variables, and

$$\ln [P/(1-P)] = C + DH + FX$$

for dichotomous (0/1) outcome variables.

These additions do not change sample sizes in any way, since all background X variables used as covariates if not observed are imputed for all cases in the analysis sample (i.e., for all children with completed spring 2003 assessments when analyzing cognitive outcomes other than the PELS and for all children with completed parent interviews when analyzing other outcomes; see Appendix 4.1 for details). Methods of estimation and significance testing involving the key parameters, B and D, which still capture the impact of access to Head Start, are also unchanged from those described in the no-covariates case.[4] Selection of the particular X variables to be

[4] With covariates added, the conversion of D from log-odds space into an estimate of impact on the probability of a positive outcome is computed using the (weighted) mean values of the new X variables across all observations in the analysis sample for a given age cohort.

included in each impact regression is discussed elsewhere in the report (including the Chapter 4 text, Appendix 4.5, and Chapters 5 through 8 of impact findings by domain).

Separate variables for fall 2002 PPVT III adapted scores are included in the impact regressions for children assessed initially in English and children assessed initially in Spanish to allow each covariate to play a distinct explanatory role in predicting spring outcomes. Vocabulary measures as indicators of child development depend heavily on the skills of the assessed child in the particular language in which the test is administered, in this case English. As a result, PPVT III adapted scores likely measure different aspects of language and literacy skills for predominantly English-speaking children and predominantly Spanish-speaking children in the fall of 2002, given their substantially different capabilities with the English language. Separate variables are needed to allow the regressions to take advantage of these distinct meanings when predicting outcomes.

In order to do this in analyses that include children with both Spanish- and English-language backgrounds in the same regression, a numeric value for the "English PPVT III adapted" variable is artificially assigned to children originally assessed in Spanish and a numeric value for the "Spanish PPVT III adapted" variable is artificially assigned to children originally assessed in English. The value 0 is used in each case, though any single common number would work. However, if included among the X variables with no further adjustments to the model, these "artificial zeroes" would distort estimates of the coefficients in the model since they do not have the same meaning as true 0s nor can they extend the linear scale followed by other real values to the 0 point on the axis without distorting how the model uses real values to predict outcomes.

We neutralized this potentially distorting effect by including in the set of X variables in the model a pair of 0/1 dummy variables that flag observations where artificial 0s have been inserted, one dummy variable for artificial 0s in the "English PPVT III adapted" variable and one dummy variable for artificial 0s in the "Spanish PPVT III adapted" variable. To see how this averts distortions of the other regression coefficients, consider how the linear models used (straight linear for continuous outcomes, linear in the log-odds ratio for categorical outcomes) would seek to accommodate the 0 values if no dummy variables are added. The first problem is that true 0s do not exist on the PPVT III adapted scale. This could be countered by inserting an artificial value that does fall within the defined range. However, that does not fully address the problem, nor is it actually necessary as opposed to using the more transparent device of inserting artificial 0s. Whatever the value chosen to "fill the gap" in these special *X* variables, its relationship to the outcome variable *Y* will have to be reflected by the same estimated e or f coefficient on PPVT III adapted scores as the influence of other real values on these variables. This clearly would create inaccuracies in how the model accounts for pre-test scores and, if those scores are correlated with other variables in the model, open the door to distortions in how the model represents those factors as well. Moreover, it will seriously diminish the amount of predictive information the model can extract from the pre-test measures, the very purpose for including them in the regressions in the first place.

A dummy variable for observations with artificial 0s in the English PPVT field and another dummy variable for observations with artificial 0s in the Spanish PPVT field remove this threat by giving the model the perfect ability to predict the average outcome of those cases from the coefficients on the dummy variables alone. If the *Y* outcome variables of these cases are distinctive at all (as we would expect), their tendency to be above or below the point at which a properly fitted linear model—the model we do not want to distort—hits the axis (the 0 point on the PPVT score number line) will be fully reflected in the coefficient on the associated dummy variable. That coefficient will provide precisely the upward or downward shift needed to account for what is distinctive about the artificial 0 cases' outcomes on average without disturbing any of the other coefficients in the model. Hence, the simultaneous addition of covariates with artificial 0s and, in each case, the corresponding "neutralizing" dummy variable will leave all the other estimates from the model unchanged, including, crucially, the estimate of Head Start's impact on that outcome, the coefficients *e* or *f*. At the same time, this paired insertion of covariates helps account for more of the variation in outcomes within the group of children who have real values in the English PPVT and Spanish PPVT fields.

An identical approach is used to take advantage of the distinctive predictive information in the fall 2002 Woodcock-Johnson III Letter-Word Identification scores of children in one instance assessed predominantly in English and, in the other instance, children assessed predominantly in Spanish. Like the PPVT III adapted, the Woodcock-Johnson III Letter-Word Identification test was administered to both sets of children in English, so that its meaning depends on the quite different English language skills of the two groups. Dual insertion of language-specific versions of this variable from the fall, together with "neutralizing" dummy variables flagging the cases where those variables contain meaningless artificial 0 values, again addresses this issue effectively.

Adding Interaction Terms

A final extension of the regression analysis interacts selected demographic characteristics and pretest measures with the Head Start group dummy variable, *H*, in order to explore if and how the intervention's impact varies among different types of children. A single regression provides information on two questions of interest, both addressed in Chapter 9 of the text: how impacts vary with the moderating factor examined and, if that factor is a 0/1 indicator of membership in a particular group, how large an impact Head Start had on each of the subgroups defined by the moderator variable (or the set of moderator variables, if a given dimension defines more than two subgroups; e.g., race/ethnicity). Letting *Z* represent the moderating variable or set of variables of interest, where *Z* may or may not have been among the covariates previously included in the regressions through *X*, the impact regression equations become

$$Y = A + BH + EX + QZ + RHZ$$

for continuous outcome variables, and

$$\ln [P/(1-P)] = C + DH + FX + SZ + THZ$$

for dichotomous (0/1) outcome variables.

A number of different coefficients in these models play important roles in addressing the questions of interest. Suppose *Y* is a continuous outcome variable, such as the number of time outs used by parents to discipline their children in the last week, and *Z* is a simple two-category dummy variable distinguishing boys ($Z=0$) from girls ($Z=1$). The logit model for dichotomous outcome variables is exactly the same for a given moderator variable—although the type of moderator will matter. In this example,

- *B*, the coefficient on the dummy variable for assignment to the Head Start sample, *H*, represents the impact of Head Start on the subgroup of children *not* flagged by the dummy variable *Z*—e.g., boys, for whom $Z=0$. For these children, the regression equation reduces to $Y = A + BH + EX$, paralleling the equation used previously to determine impacts on all children.
- *Q*, the coefficient on the moderator variable *Z*, shows how much higher or lower average outcomes are for girls ($Z=1$) than for boys ($Z=0$), when children do *not* have access to Head Start—that is, for children in the non-Head Start sample for whom $H=0$. For these children, the regression model simplifies to $Y = A + EX + QZ$, highlighting the role of *Z* in influencing spring 2003 outcomes but not telling us anything about the impact of Head Start.
- *R*, the coefficient on the interaction of the Head Start sample dummy variable and the moderator variable, *HZ*, indicates the difference between the average impact of the

intervention on girls ($Z=1$) and the average impact on boys ($Z=0$), the latter previously identified as B. From this we infer that

- $B+R$ is the average impact of Head Start on girls, the $Z=1$ group. For these children, the model simplifies to $Y = A + BH + EX + Q + RH$ or, rearranging terms, $Y = [A + Q] + [B+R]H + EX$, again paralleling in terms of the intercept and H and X explanatory variables the equation previously used to determine impacts on all children.

This way of looking at the results—once $A, B, E, Q,$ and R are estimated as $a, b, e, q,$ and r — highlights impacts on subgroups: b for the $Z=0$ group, $b+r$ for the $Z=1$ group. Statistical significance tests on this coefficient and sum (i.e., linear combination) of coefficients tell us whether either impact differs significantly from 0, using test procedures available for this purpose in OLS (for continuous outcomes) and logit (for dichotomous outcomes).[5]

Another perspective on the same set of regression coefficients and related tests, highlights Z's role as a moderator of the size of impact without reference to average impact on any particular subgroup. Restating the third bullet above as

- R, the coefficient on the interaction of the Head Start sample dummy variable and the moderator variable, HZ, indicates *the degree to which Z (in this case gender) alters the size of Head Start's impact.*

This perspective shines through when the complete regression equation is reordered and certain terms taken apart and regrouped:

$$Y = A + EX + QZ + [B + RZ]\,H\,.$$

This formulation emphasizes how the size of Head Start's influence— $[B + RZ]$ — may vary with the background variable Z; i.e., how Z may moderate the impact of the intervention to create differences in the degree to which different types of children benefit from the program. The estimate of R and its test of significance measure the existence and strength of this moderating influence.

Continuous moderator variables, such as maternal depression scores or the pretest values of a cognitive assessment, fit easily into this last formulation. If Z can take on a range of values over an expansive ordinal and cardinal scale, no one value of Z carves out a subgroup of Head Start participants for special focus and an exclusive impact estimate. Rather, the main question is

[5] Actual computation of impact estimates for different subgroups requires conversion from log-odds space to probability units for the dichotomous outcome variables. For this purpose, the covariates in the model (other than Z) are set to their (weighted) mean values for the entire analysis sample in a given age cohort. The moderator Z is set first to 0 and probabilities with H equal to 1 (for the Head Start group) and H equal to 0 differenced to get the impact estimate for the $Z=0$ subgroup. Then Z is set to 1 and probabilities with H equal to 1 (for the Head Start group) and H equal to 0 differenced to get the impact estimate for the $Z=1$ subgroup.

whether this factor exerts an influence on the size of impact across its entire range. The $[B + RZ] H$ term in the preceding equation conveys a linearized version of this influence: an impact may occur when children gain access to Head Start ($H=1$, rather than $H=0$) and, if so, that impact may vary in size with the moderating factor Z along the slope and intercept defined by $B + RZ$.[6] Thus, the test of the statistical significance of r, our estimate of R, (or of t, the estimate of T in the logit version of the model for a dichotomous outcome variable) and consideration of its magnitude tells us about the influence of Z as a moderator whether it be a categorical dummy variable distinguishing one subgroup of children from another or a continuous measure describing the characteristics of a whole range of individual children.

A final variation of the moderator/subgroup analysis occurs by replacing Z with a *collection* of two or more categorical dummy variables. In the current report, this arises only when looking at the moderating influence of race/ethnicity and at subgroup impacts for children in different racial/ethnic groups. Here, we use two Z variables, call them Zh and Zb, which flag Hispanic and non-Hispanic Black children respectively. The OLS version of the regression equation in this instance expands slightly into

$$Y = A + BH + EX + Q[Zh] + RH[Zh] + U[Zb] + VH[Zb]$$

Here we have the standard subgroup model with the main effect of the moderator and its interaction with the intervention echoed twice at the end of the equation, once for Hispanic children — $Q[Zh] + RH[Zh]$ — and once for non-Hispanic Black children — $U[Zb] + VH[Zb]$. As before, the coefficient on H alone (i.e, B) gives the impact of Head Start on the $Z=0$ subgroup, only here that subgroup is defined by *all* Z moderator variables in the model equaling 0: $Zh = Zb = 0$. In other words, b, our estimate of B, measures the impact of the program on non-Hispanic, non-Black children.

This point echoes that made first in the original set of four bullet points above—how the model captures the impact of Head Start on the "omitted" group. The interpretations of estimated coefficients in the other three original bullets also apply here, repeated twice, once for Hispanic

---

[6] Converted from log-odds space to probability units, the size of impact on a dichotomous outcome variable will vary with Z in a nonlinear fashion that depends on the values of the other covariates in the model (the *X*s) and the precise level of Z itself (through the SZ term). As always, the *X*s are set to their (weighted) mean values for the entire analysis sample in a given age cohort. For parallelism, the measure of variation in impact with changes in Z (e.g., maternal depression) is then calculated at the sample-wide (weighted) mean value of Z, recognizing that this is a "local" slope coefficient showing how impact on probability varies in size with the moderating factor in the near vicinity of Z's mean value and may not apply elsewhere in the range of Z values.

children and once for Black children. Most important in analyzing subgroup impacts are the two versions of the final bullet in that collection:

- $B+R$ is the average impact of Head Start on Hispanic children (the $Zh=1$, $Zb=0$ subgroup). For these children, the model simplifies to $Y = A + BH + EX + Q + RH$ or, rearranging terms, $Y = [A + Q] + [B+R]H + EX$, paralleling the equation originally used to determine impacts on all children.
- $B+V$ is the average impact of Head Start on Black children (the $Zh=0$, $Zb=1$ subgroup). For these children, the model simplifies to $Y = A + BH + EX + U + VH$ or, rearranging terms, $Y = [A + U] + [B+V]H + EX$, again paralleling the equation originally used to determine impacts on all children.

These—plus the just preceding interpretation of B—are the basis for the reported magnitudes and tests of the statistical significance Head Start's impact on the three racial/ethnic subgroups.

Appendix 4.4: Measures of Fall 2002 "Starting Points" Used in the Regression Models, By Child and Parent Outcomes

Outcome Measure	Fall 2002 Measure Used As a Covariate, by Language Group
Cognitive Domain	
Peabody Picture Vocabulary Test	PPVT
Comprehensive Test of Phonological and Print Processing (CTOPPP), Elision Subtest	*Assessed Primarily in English in Fall 2002* Comprehensive Test of Phonological and Print Processing (CTOPPP), Elision Subtest *Assessed Primarily in Spanish in Fall 2002* PPVT *All Language Groups Combined* PPVT
Letter Naming Task	PPVT
Color Naming	Color Naming
Counting Bears	Counting Bears
McCarthy Scales of Children's Abilities Draw-a-Design Subtest	McCarthy Scales of Children's Abilities Draw-a-Design Subtest
Woodcock-Johnson III: Letter-Word Identification	Woodcock-Johnson III: Letter-Word Identification
Woodcock-Johnson III: Spelling	*Assessed Primarily in English in Fall 2002* Woodcock-Johnson III: Spelling *Assessed Primarily in Spanish in Fall 2002* PPVT *All Language Groups Combined* PPVT
Woodcock-Johnson III: Applied Problems	*Assessed Primarily in English in Fall 2002* Woodcock-Johnson III: Applied Problems *Assessed Primarily in Spanish in Fall 2002* PPVT *All Language Groups Combined* PPVT
Woodcock-Johnson III: Oral Comprehension	*Assessed Primarily in English in Fall 2002* Woodcock-Johnson III: Oral Comprehension *Assessed Primarily in Spanish in Fall 2002* PPVT *All Language Groups Combined* PPVT
Test de Vocabulario en Imágenes Peabody (TVIP)	*Assessed Primarily in Spanish in Fall 2002* Test de Vocabulario en Imágenes Peabody (TVIP)

Outcome Measure	Fall 2002 Measure Used As a Covariate, by Language Group
Cognitive Domain	
Batería Woodcock-Muñoz Pruebas de aprovechamiento-Revisada: Identificación de letras y palabras	*Assessed Primarily in Spanish in Fall 2002* Batería Woodcock-Muñoz Pruebas de aprovechamiento-Revisada: Identificación de letras y palabras
Parent (reported) Emergent Literacy Scale (PELS)	Parent (reported) Emergent Literacy Scale (PELS)
Social-Emotional Domain	
Social Skills and Positive Approaches to Learning	Social Skills and Positive Approaches to Learning
Total Child Behavior Problems Scale	Total Child Behavior Problems Scale
Aggressive Behavior Scale	Aggressive Behavior Scale
Hyperactive Behavior Scale	Hyperactive Behavior Scale
Withdrawn Behavior Scale	Withdrawn Behavior Scale
Social Competencies Checklist	Social Competencies Checklist
Parenting Practices Domain	
Parent used time out in the last week	Parent used time out in the last week
Number of times parent used time out in the last week	Number of times parent used time out in the last week
Parent spanked child in the last week	Parent spanked child in the last week
Number of times parent spanked child in the last week	Number of times parent spanked child in the last week
Parental Safety Practices Scale	Parental Safety Practices Scale
Removing Harmful Objects Scale	Removing Harmful Objects Scale
Restricting Child Movement Scale	Restricting Child Movement Scale
Safety Devices Scale	Safety Devices Scale
Family Cultural Enrichment Scale	Family Cultural Enrichment Scale
How many times child was read to in the last week by parent or other family member	How many times child was read to in the last week by parent or other family member
Health Domain	
Child seen by dentist since last September	Child seen by dentist since last September
Overall child’s health status	Overall child’s health status
Child has injury in last month requiring medical treatment	Child has injury in last month requiring medical treatment
Child has health insurance	Child has health insurance
Child has condition that requires ongoing medical care	Child has condition that requires ongoing medical care

Appendix 4.5: Tests for Lack of Impact of Head Start on Demographic and Developmental Factors Measured in Fall 2002

As discussed in Chapter 4, most of the demographic and developmental factors considered as covariates or for the impact regressions were measured with some lag following random assignment. Because these measures could have been influenced by the Head Start intervention, a statistical procedure was developed to determine if strong evidence exists that a given measure **was not affected** by Head Start to any appreciable degree. Where such evidence was found, the measure remained in the regression equation for the preferred impact analysis and, if of substantive interest, served as a moderator in examining how impacts varied with family and child background characteristics. This appendix describes the procedure used to make this determination and presents test results for all fall 2002 variables tested using the technique.

The procedure adopted seeks a 90 percent assurance that Head Start's impact in the fall was small or nonexistent. Only then is it considered safe to rely on impact estimates that adjust for a given background characteristic of families or children in the study or the fall measure of the outcome variable. "Small" is defined on a relative basis that takes account of how much the fall measure varies in the population being studied, as indicated by the standard deviation of the outcome variable in fall 2002 for the non-treated comparison group. A guideline suggested by Cohen is adopted that classifies the impacts of educational and child development interventions as small, modest, or large based on their "effect size"—their average impact divided by the standard deviation in the population.[1] If 90 percent certain that effect size is less than 0.2, we conclude that a small (or perhaps zero) impact has taken place, making it safe to include the variable in question in impact regressions as a covariate and/or moderator. Otherwise, where true impact may move into the range characterized by Cohen as "modest" or larger, the analysis omits the variable from the preferred set of covariates and refrains from using it as a moderator.

Formally, the test involves constructing a 90 percent confidence interval for true impact in fall 2002 calibrated in effect-size units, then checking that this interval lies entirely between -0.20 (the limit of negative impacts that Cohen would consider "small") and 0.20 (the limit of positive impacts that Cohen would consider "small"). First, the regression procedures described in Appendix 4.3 in connection with the main impact findings in spring 2003 are applied using the

---

[1] See Cohen, J. (1987). ***Statistical Power Analysis for the Behavioral Sciences*** (2nd ed.). Hillsdale, NJ: Erlbaum.

fall 2002 measures of interest as dependent variables. Thus, we estimate impacts on 32 continuous variables using OLS regression models that express a fall measure, F, as the sum of an intercept term (a) and a **shift in the intercept produced by a dummy variable for inclusion in the Head Start group (H):**

$$F = a + bH$$

This equation is estimated using weighted data described in Appendix 1.2 to represent the national population of newly entering Head Start children in communities with more potential Head Start participants than funded Federal Head Start slots.

The coefficient on the dummy variable in this model, b, provides an estimate of B, the true (but ultimately unknown) average impact of Head Start on the fall measure F—an estimate computationally identical to the difference between the average level of the fall measure for the Head Start sample (H=1) and the average level of the measure for the non-Head Start sample (H=0). The estimate of the standard deviation of this coefficient—its standard error, s—is used to construct an interval around b that for 90 percent of all possible surveys conducted in the same manner will contain the true average impact B, given the usual OLS assumption that the dependent variable and hence all estimated coefficients have normal distributions. This interval is

$$b - 1.645s < B < b + 1.645s, \text{ since}$$

$$\begin{aligned} Pr(b - 1.645s < B < b + 1.645s) &= Pr[(b-B)/s - 1.645 < 0 < (b-B)/s + 1.645) \\ &= Pr[-1.645 < (b-B)/s < 1.645] = M(1.645) - M(-1.645) = 0.95 - 0.05 \\ &= 0.90, \end{aligned}$$

where M( ) is the cumulative density function of a standard normal distribution used to approximate the students t distribution of (b-B)/s when sample size is large (e.g., N > 200). The endpoints of this interval are then divided by the standard deviation of F in the non-Head Start sample, d,[2] to convert it to effect-size units:

$$[b - 1.645s] / d < B/d < [b + 1.645s] / d.$$

If $\quad -0.20 < [b - 1.645s] / d \quad \text{and} \quad [b + 1.645s] / d < 0.20,$

we conclude that the true effect of Head Start on F measured in fall 2002, B/d, is within Cohen's range of "small". By construction, there is at least a 90 percent chance that we are right in this

[2] Standard deviation is calculated as the square root of the sum of squared deviations of individual values of *F* from the mean of *F* for the entire non-Head Start sample.

conclusion, since the entire 90-percent confidence interval for true effect size is between -0.20 and 0.20. If instead either

$$[b - 1.645s] / d \; < \; -0.20 \qquad or \qquad [b + 1.645s] / d \; > \; 0.20$$

we cannot be 90 percent confident that the true effect size is "small" by Cohen's standard.

We proceed in parallel fashion for the 21 0/1 indicator variables, but use logistic regression procedures described in Appendix 4.3 to calculate a log-odds ratio and obtain an estimate of *B* and its standard error, *s*. The estimated equation expresses the natural log of the odds ratio of $F=1$ to $F=0$ as the sum of an intercept term and a shift in the intercept produced by a dummy variable for inclusion in the Head Start group:

$$\ln [P/(1-P)] \; = \; c + dH,$$

where P *is the probability that* F = 1 *(and, hence,* 1-P *is the probability* F = 0*).*

The impact estimate *b* is then derived as the difference between two quantities:

- log-odds ratio for children in the Head Start group, $c + dH$, converted into a probability by passing it through the logistic transformation,

  $P(H=1) = exp\,(c+d) / [1 + exp\,(c+d)]$, and

- the log-odds ratio for children in the non-Head Start group, *c*, converted into a probability by the same transformation,

  $P(H=0) \; = \; exp\,(c) / [1 + exp\,(c)]$.

Thus,

$$b \; = \; exp\,(c+dH) / [1 + exp\,(c+dH)] \; - \; exp\,(c) / [1 + exp\,(c)] \,.$$

The standard error of *b*, *s*, and a 90 percent confidence interval for *B* follow from the standard probability and distributional assumptions of logit regressions, as calculated by the SUDAAN software package.

Appendix 4.6: Basis for Assuming That Non-Participants Experienced No Intervention Effects

Although a common assumption in experimental evaluations of social program impacts, the decision to consider nonparticipants from the Head Start group unaffected by the intervention merits discussion. The validity of this assumption begins with the very exacting definition of a "nonparticipant" that has been adopted for this study, i.e., if a child attended a Federal Head Start program for even a day during the 2002-2003 program year, she or he was considered a "participant." This means that while the chances are slim that such a child (or her/his family) could be meaningfully influenced by so brief an experience, this definition allows that some Head Start impact could have occurred in this case. Only those children identified as *never* having attended Head Start were considered nonparticipants and made part of the "no-show" adjustment. In addition, the set of Head Start group members considered no-shows was further narrowed—thus strengthening the case for assuming zero impacts on nonparticipants even more—by looking at multiple sources of information for indications of participation as was described in Chapter 2 and in Appendix 2.2. Evidence of Head Start attendance from any of these sources took a child out of the "no-show" group and released him/her from the assumption of no program impact.

With such careful reduction of the non-participant group to only those children who never spent any time in any Federal Head Start program, only one means of being affected by the program remains open to those individuals—the possibility that just **applying to Head Start and going through the random assignment and admission process** influenced behavior and outcomes. Before this could create impacts on nonparticipants that would show up in the experimental comparison outcomes between the two groups created at random assignment, it would have to occur for children randomized into the group accepted into the program (but who never attend) **and not to children randomized into the non-Head Start group.** This follows from the fact that any symmetric impact of the application and randomization process on both groups will net out in the calculation of the experimental impacts, which focuses solely on the **difference** in outcomes between the Head Start and non-Head Start groups. The application and random assignment process was identical up to the point when program staff were told which children had been selected into the Head Start group; during this interval, when no one (not even study staff) knew who would be selected into the program and who would not be, it was impossible for different, non-canceling impacts to have occurred.

This leaves only the interval following notification of the program staff of admission of no-shows into the program as a time when differential impacts could have occurred. There were two possibilities during this period for the existence of the Head Start program to influence a nonparticipating child's (or his/her family's) outcomes:

- Head Start program staff might do something to guide families to alternative non-Head Start services that those families would not have accessed otherwise, and those services influenced later child and family outcomes; and/or

- The families themselves might choose different non-Head Start options, or alter the timing of when they used those options, simply because they knew Head Start was available to their child—and the use of those services then later alters child and family outcomes.

The potential for either of these influences to appreciably contribute to the overall gain caused by Head Start—the quantity the no-show adjustment attributes solely to actual program participants—is remote. Encouragement, information, and assistance in using **non-Head Start** services should rarely be a priority for Head Start programs seeking to meet Federal enrollment targets and that have already expressed the intention of (and been given "permission" by the random assignment process to) involving the child in Head Start itself. Moreover, there was some chance that similar children in the non-Head Start sample received this kind of aid as well,[1] which would then result in a canceling out of its effects in the impact analysis. The window for parents to behave differently because Head Start admission had been granted but not pursued was presumably short, assuming that parents understand the offered "Head Start slot" would not be held open for their child indefinitely. Once that slot was presumed by the family to have been filled by some other child, the fact that it was once open to the family's child should have no further effect on behavior.

The one real opportunity for Head Start to affect the longer run outcomes of no-shows arises in cases where notification of admission reached families well ahead of the time their children could actually begin their Head Start participation. Many grantees and delegate agencies focus new admissions on the start of the school year and also seek to inform families that their child has been admitted for the fall in the late spring or summer. The random assignment process was set up to accommodate this practice as much as possible, which meant that many families

[1] Agreements with grantees and delegate agencies not to serve children and families assigned to the control group did not preclude referral for other services—or even direct provision of those services if no Federal Head Start funding was involved—for grantees/delegate agencies whose contacts with the community through its non-Head Start activities are extensive and often involve information and advice on service options. These activities, and their long-run consequences, are considered outside the Head Start intervention per se as something that would occur even if Federal funding for Head Start did not exist.

randomized into the Head Start group found out about this decision a number of weeks, or even months, before participation could begin. These families may have behaved differently in the interval between that point and when they presumed their "unclaimed" Head Start slot had been relinquished to another child (i.e., differently from families assigned to the non-Head Start group who were told they did *not* have option of joining the program). This behavior could then have led to different outcomes the following spring, making "no-shows" to some extent contributors to the total gain caused by the program.

The most likely type of altered family behavior to emerge during the "waiting period" between Head Start admission and the start of program services in the fall was a **reduction** in the pursuit of other types of child development assistance. Parents expecting to rely on Head Start to support their child's development in fall 2002 presumably had less incentive to find alternative supports during the summer and to arrange in advance non-Head Start services for their child. As a consequence, their children's outcomes could be set back relative to counterparts in the non-Head Start sample. This did not happen for families that quickly decided their children would not attend Head Start in the fall even though admitted (e.g., those who moved at the end of the previous school year), but this was the minority of all eventual no-shows. It is also possible that some families did **more** to push forward their children's development while waiting for Head Start to begin in the fall, wanting to make sure the child was ready for the new experience or inspired by the theme of Head Start—of which they (temporarily) considered themselves a part—emphasizing intellectual stimulation and engagement of the child.

Overall, it is hard to gauge how much, and in what direction, the "anticipatory effects" of Head Start participation altered family behaviors over the summer for those families that ultimately did not participate. But, to assume there was an effect requires the following to be true:

- Such families behaved differently;
- Children ended up in different places cognitively and in terms of behavior and health care the following spring (when outcome data were colleted) as a result; and
- This contribution to the overall measured gain attributable to Head Start appreciably affected the size of the measured average impact of the program among all children assigned to the Head Start group and, hence, the size of the average impact on participants inferred through the no-show adjustment.

However, because one cannot confidently rule this out completely, it is necessary to acknowledge that some small bias—either up or down—may be present in the no-show adjusted impact estimates for participants presented in Appendices 5.1, 6.1, 7.1, and 8.1.

Appendix 5.1: Cognitive Domain, Estimated Impact on Program Participants

This appendix presents the impact estimates that are adjusted for the fact that some of the children assigned to Head Start failed to take advantage of this opportunity, i.e., they never participated in Head Start (the "no-shows") and that some children assigned to the non-Head Start group managed to find their way into the program (the "crossovers"). The impact of admission into the program measures what grantees have the power to do—provide access to the program. However, the question of how much children gain from actually participating in Head Start remains an important one.

These results are summarized in Exhibit A.5.1.1 and A.5.1.2[1] (for children from the 3- and 4-year-old groups, respectively) focusing only on the statistically significant impact estimates, both overall and for the subgroup/moderator analyses. The first three columns of figures provide the impact estimates, and the last three columns provide the effect sizes associated with each estimate. Three different impact estimates are provided in the tables: (1) the impact of **access** to Head Start (the intent-to-treat estimates discussed in Chapter 5); (2) the impact of Head Start **participation,** adjusting only for the occurrence of "no shows"; and, (3) the impact of Head Start **participation,** adjusting for the occurrence of **both** "no shows" and "crossovers" (these are shown only for the overall average impacts). The latter two estimates, impacts on those who actually receive Head Start services, are often called "impact on the treated."

The last three columns express each of the respective estimates as "effect sizes," which are defined as the impact estimates divided by the standard deviation of the outcome measure in the population, providing a "yardstick" for gauging the quantitative importance of a measure impact in relation to the natural variation of the child or family outcome Head Start is seeking to effect.[2] Effect sizes are important in interpreting the size of Head Start's measured impact and, in particular, how much larger that impact may be for the average program participant as opposed to the larger group of children and families accorded access to the program.

---

[1] See Chapter 4 for an explanation of the terms used here and the analytical methods used to make the statistical adjustments.

[2] The standard deviation of each outcome measure is derived from data on children/families in the non-Head Start sample, excluding members of the Head Start sample, to ensure that measures of underlying variation are not affected by the intervention.

As can be seen, all measures of the impact of *participation* in Head Start exceed the corresponding measures of the average impact of *access to* Head Start shown in the first column, regardless of which adjustment method is used. For example, the first row of Exhibit A.5.1.1 shows an estimated impact of 4.23 scale points on the Peabody Picture Vocabulary Test (PPVT) for the average child in the 3-year-old group granted access to the program through assignment to the Head Start research sample and a larger average impact of 4.50-4.79 scale points for those children who actually attended Head Start. These same increases are seen in the relationship between the calculated effect sizes for the estimates of "access" vs. "participation," with some instances of the size of the effect moving from "small" (under 0.2) to "modest (between 0.2 and 0.5).

These tables also show the difference in the magnitude of the adjustment associated with the no-show correction alone and the combined no-show and crossover adjustment. The differences between the intent-to-treat estimates and the no-show/crossover adjusted estimates are, in most cases, larger than the differences shown using just the no-show correction alone. This is a clear indication of the dampening effect of the crossover children in the non-Head Start group, i.e., the fact that these children were able to receive some Head Start program services reduces the size of any observed difference in outcomes between the Head Start and non-Head Start groups. However, although these data represent our best estimate of this phenomenon, a more thorough analysis is needed and will be addressed in future reports. In the meantime, crossover-adjusted estimates have only been calculated for the overall average effects. Consequently, the estimates incorporating the combined no-show/crossover adjustment should only be used as a rough indication of the likely consequences of the presence of crossovers on the interpretation of the impact of Head Start on children's school readiness.

Exhibit A.5.1.1: Initial Estimates of the Impact of Head Start on Cognitive Outcomes, Intent to Treat, and Impact on the Treated, Statistically Significant Results Only, 3-Year-Old Group, Combined English-English and Spanish-English Group (Weighted Data)

Outcome Measure	Impact Estimates			Effect Sizes		
	Impact of Access to Head Start	**Impact of Head Start Participation, No-Show Adjustment**	**Impact of Head Start Participation, Combined No-Show and Crossover Adjustment**	**Impact of Access to Head Start**	**Impact of Head Start Participation, No-Show Adjustment**	**Impact of Head Start Participation, Combined No-Show and Crossover Adjustment**
Overall Impact						
PPVT-III	4.23*	4.79*	4.50*	0.12	0.14	0.13
WJ-III Letter-Word Identification—IRT/ML	5.65**	6.40**	8.05***	0.24	0.27	0.34
Letter Naming Task	1.30*	1.47*	2.05***	0.19	0.22	0.31
Color Naming/Identification	0.70*	0.79*	0.77	0.10	0.11	0.11
McCarthy Drawing	0.15*	0.17*	0.23**	0.13	0.15	0.20
PELS	0.47***	0.53***	0.65***	0.34	0.38	0.47
Difference in Impact [1]						
PPVT-III: Depression	-0.11*	-0.12*	N/A	- 0.00	- 0.00	N/A
CTOPPP Elision: Depression	-0.13*	-0.15*	N/A	- 0.00	- 0.00	N/A
WJ-III Applied Problems: Depression	-0.09*	-0.10*	N/A	- 0.00	- 0.00	N/A
WJ-III Oral Comprehension: Race (White Impact Exceeds African American)	4.73*	5.36*	N/A	0.33	0.38	N/A
Color Naming/Identification: Depression	-0.02*	-0.02*	N/A	- 0.00	- 0.00	N/A
Counting Bears: Depression	-0.00*	-0.00*	N/A	- 0.00	- 0.00	N/A
PELS: Depression	-0.00*	-0.00*	N/A	- 0.00	- 0.00	N/A

Exhibit A.5.1.1: (continued)

Outcome Measure	**Impact Estimates**			**Effect Sizes**		
	Impact of Access to Head Start	**Impact of Head Start Participation, No-Show Adjustment**	**Impact of Head Start Participation, Combined No-Show and Crossover Adjustment**	**Impact of Access to Head Start**	**Impact of Head Start Participation, No-Show Adjustment**	**Impact of Head Start Participation, Combined No-Show and Crossover Adjustment**
Impact on Subgroup [2]						
PPVT-III: Parent Married	4.51*	5.20*	N/A	0.13	0.15	N/A
PPVT-III: Spanish-English Language Group	7.52*	8.54*	N/A	0.21	0.24	N/A
PPVT-III: Hispanic	7.26*	8.24*	N/A	0.21	0.23	N/A
CTOPPP Elision: African American	7.47*	8.40*	N/A	0.17	0.18	N/A
WJ-III Letter-Word Identification: No Special Needs	5.38**	6.01**	N/A	0.22	0.25	N/A
WJ-III Letter-Word Identification: African American	5.80**	6.52**	N/A	0.24	0.27	N/A
WJ-III Letter-Word Identification: Hispanic	6.92*	7.85*	N/A	0.29	0.33	N/A
WJ-III Letter-Word Identification: English-English Language Group	5.05***	5.72***	N/A	0.21	0.24	N/A
WJ-III Letter-Word Identification: Parent Married	6.53**	7.52**	N/A	0.27	0.31	N/A
WJ-III Letter-Word Identification: Parent Not Married	5.21**	5.82**	N/A	0.22	0.24	N/A
WJ-III Spelling: Hispanic	5.61*	6.37*	N/A	0.25	0.28	N/A
Letter Naming Task: No Special Needs	1.24*	1.41*	N/A	0.19	0.21	N/A
Letter Naming Task: Hispanic	1.45*	1.65*	N/A	0.22	0.25	N/A
Letter Naming Task: Parent Not Married	1.46*	1.63*	N/A	0.22	0.25	N/A
WJ-III Oral Comprehension: White	2.82**	3.24**	N/A	0.20	0.23	N/A

Exhibit A.5.1.1: (continued)

Outcome Measure	**Impact Estimates**			**Effect Sizes**		
	Impact of Access to Head Start	**Impact of Head Start Participation, No-Show Adjustment**	**Impact of Head Start Participation, Combined No-Show and Crossover Adjustment**	**Impact of Access to Head Start**	**Impact of Head Start Participation, No-Show Adjustment**	**Impact of Head Start Participation, Combined No-Show and Crossover Adjustment**
WJ-III Oral Comprehension: Parent Married	2.09*	2.41*	N/A	0.15	0.17	N/A
Color Identification: English - English Language Group	0.87**	0.99**	N/A	0.12	0.14	N/A
Color Identification: Parent Married	1.50**	1.73**	N/A	0.21	0.24	N/A
McCarthy Drawing: No Special Needs	0.16*	0.18*	N/A	0.14	0.16	N/A
McCarthy Drawing: African American	0.18*	0.20*	N/A	0.16	0.17	N/A
McCarthy Drawing: English-English Language Group	0.17*	0.19*	N/A	0.15	0.17	N/A
PELS: No Special Needs	0.50***	0.57***	N/A	0.36	0.41	N/A
PELS: White	0.37***	0.43***	N/A	0.27	0.31	N/A
PELS: African American	0.53**	0.60**	N/A	0.38	0.43	N/A
PELS: Hispanic	0.51**	0.58**	N/A	0.37	0.42	N/A
PELS: English-English Language Group	0.48***	0.54**	N/A	0.35	0.39	N/A
PELS: Spanish-English Language Group	0.46*	0.52*	N/A	0.33	0.38	N/A
PELS: Parent Married	0.52***	0.60***	N/A	0.38	0.43	N/A
PELS: Parent Not Married	0.43***	0.48***	N/A	0.31	0.35	N/A

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.
[1] A total of 82 differences in impacts between subgroups were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 5.2. Findings for depression indicate the change in Head Start's estimated impact that accompanies a 1-point increase in mother's baseline depression score. Findings for baseline factors other than depression indicate the amount by which Head Start's estimated impact for the first subset of participants listed in the row label exceeds that for the second subset listed.
[2] A total of 99 subgroup impacts were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 5.2.

Exhibit A.5.1.2: Initial Estimates of the Impact of Head Start on Cognitive Outcomes, Intent to Treat, and Impact on the Treated, Statistically Significant Results Only, 4-Year-Old Group, Combined English-English Spanish-English Group (Weighted Data)

Outcome Measure	Impact Estimates			Effect Sizes		
	Impact of Access to Head Start	Impact of Head Start Participation, No-Show Adjustment	Impact of Head Start Participation, Combined No-Show and Crossover Adjustment	Impact of Access to Head Start	Impact of Head Start Participation, No-Show Adjustment	Impact of Head Start Participation, Combined No-Show and Crossover Adjustment
Overall Impact						
WJ-III Letter-Word Identification	5.74*	6.93*	6.83*	0.22	0.26	0.26
Letter Naming Task	2.28**	2.75**	2.39*	0.24	0.29	0.25
WJ-III-Spelling	4.14*	4.99*	4.74*	0.16	0.19	0.18
Color Identification	0.60	0.72	0.97*	0.11	0.13	0.17
PELS	0.41***	0.49***	0.52***	0.29	0.35	0.37
Difference in Impact [1]						
Counting Bears: Race (Hispanic Impact Exceeds White)	0.52*	0.63*	N/A	0.38	0.46	N/A
Impact on Subgroup [2]						
PPVT-III: Hispanic	5.64*	6.42*	N/A	0.14	0.16	N/A
WJ-III Letter-Word Identification: No Special Needs	5.88*	7.16*	N/A	0.22	0.27	N/A
WJ-III Letter-Word Identification: African American	10.56*	14.01*	N/A	0.40	0.53	N/A
WJ-III Letter-Word Identification: English-English Language Group	7.32*	9.30*	N/A	0.27	0.35	N/A
WJ-III Letter-Word Identification: Parent Not Married	7.92*	9.45*	N/A	0.30	0.35	N/A
Letter Naming Task: No Special Needs	2.39**	2.91**	N/A	0.25	0.31	N/A
Letter Naming Task: White	2.77**	3.23**	N/A	0.29	0.34	N/A
Letter Naming Task: English-English Language Group	3.05**	3.87**	N/A	0.32	0.41	N/A
Letter Naming Task: Parent Not Married	2.70*	3.22*	N/A	0.29	0.34	N/A

Exhibit A.5.1.2: (continued)

Outcome Measure	Impact Estimates			Effect Sizes		
	Impact of Access to Head Start	Impact of Head Start Participation, No-Show Adjustment	Impact of Head Start Participation, Combined No-Show and Crossover Adjustment	Impact of Access to Head Start	Impact of Head Start Participation, No-Show Adjustment	Impact of Head Start Participation, Combined No-Show and Crossover Adjustment
McCarthy Drawing: Parent Not Married	0.39*	0.47*	N/A	0.20	0.24	N/A
WJ-III Spelling: No Special Needs	4.97**	6.05**	N/A	0.19	0.24	N/A
WJ-III Spelling: African American	9.75**	12.94**	N/A	0.38	0.50	N/A
WJ-IIII Spelling: English-English Language Group	4.49**	5.70**	N/A	0.17	0.22	N/A
WJ-III Spelling: Parent Not Married	6.31*	7.53*	N/A	0.25	0.29	N/A
PELS: No Special Needs	0.43***	0.52***	N/A	0.30	0.37	N/A
PELS: African American	0.75**	0.98**	N/A	0.53	0.70	N/A
PELS: English-English Language Group	0.45***	0.57***	N/A	0.32	0.40	N/A
PELS: Parent Married	0.35*	0.43*	N/A	0.25	0.30	N/A
PELS: Parent Not Married	0.52**	0.61*	N/A	0.37	0.43	N/A

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

[1] A total of 82 differences in impacts between subgroups were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 5.2. Findings for depression indicate the change in Head Start's estimated impact that accompanies a 1-point increase in mother's baseline depression score. Findings for baseline factors other than depression indicate the amount by which Head Start's estimated impact for the first subset of participants listed in the row label exceeds that for the second subset listed.

[2] A total of 99 subgroup impacts were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 5.2

Appendix 5.2: Factors That Moderate the Impact of Head Start: Detailed Tables for Cognitive Outcomes

The following tables (Exhibits A.5.2.1 through A.5.2.22) provide the results of the moderator/subgroup analyses for measures of cognitive outcomes, with a separate table for each individual measure. For clarity, these results are only presented for the full combined sample (i.e., not separately for the English-English and Spanish-English language groups). Each table is organized as follows:

- The first column lists the variable used in the particular subgroup/moderator analysis (separate regressions were estimated for each moderator). For example, analyses were conducted to examine the extent to which child gender was related to program impact.
- As shown, separate lines are shown for the overall construct (e.g., gender) and for each of the subgroups that make up the construct, e.g., boys and girls. Estimates associated with the overall construct represent estimated **differences** in impacts (e.g., boys versus girls), while the figures associated with each of the subgroup rows represent the **impact on the individual subgroups** (e.g., impacts on boys alone).
- For comparison purposes, the next column provides the mean on the particular outcome measure for the group indicated among children in the non-Head Start group in spring 2003 (the end of the first program year).
- The next set of columns provides the estimated impact on the individual subgroups, while the last two columns provide the estimated difference in impact between subgroups.
- As with the overall impact tables provided in Chapter 5, the estimated impacts are shown using two separate estimation specifications: (1) using regression analyses that include only demographic covariates measured in fall 2002 and (2) using regression analyses that added a measure of the outcome variable assessed in fall 2002. The highlighting indicates which estimate is considered the "best" (see Chapter 4) and which is highlighted in the discussion in Chapter 5.

Exhibit A.5.2.1: Initial One Year Estimates of Factors That Moderate Head Start's Impact on the PPVT-III: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

PPVT-III					
Moderator/Subgroup (Sample N=2,071)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				4.19	3.78
Special Needs	247.16	7.07	7.55		
No Special Needs	250.34	2.88	3.77		
Child's Race					
White	262.51	1.48	3.34	0.20 (vs. Black)	1.15 (vs. Black)
Black	247.10	1.68	2.19	5.41 (vs. Hispanic)	5.07 (vs. Hispanic)
Hispanic	239.51	7.09	7.26*	5.61 (vs. White)	3.92 (vs. White)
Caregiver Depression (Continuous)				-0.12*	-0.11*
Language of Assessment				4.49	4.10
English-English	255.52	2.43	3.42		
Spanish-English	225.12	6.92	7.52*		
Fall PPVT					0.02
Parent Married				0.91	0.53
Married	252.54	4.15	4.51*		
Not Married	247.85	3.25	3.98		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.5.2.2: Initial One Year Estimates of Factors That Moderate Head Start's Impact on the PPVT-III: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

PPVT-III					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				2.33	0.64
Special Needs	287.76	4.06	3.15		
No Special Needs	291.79	1.73	2.51		
Child's Race					
White	312.72	2.16	0.96	3.29 (vs. Black)	1.12 (vs. Black)
Black	294.60	-1.14	-0.16	4.97 (vs. Hispanic)	5.80 (vs. Hispanic)
Hispanic	272.69	3.84	5.64*	1.68 (vs. White)	4.68 (vs. White)
Caregiver Depression (Continuous)				0.01	0.00
Language of Assessment				4.61	3.60
English-English	303.58	1.03	1.50		
Spanish-English	262.52	5.64*	5.09		
Fall PPVT					-0.01
Parent Married				3.53	3.54
Married	290.22	4.65	4.97		
Not Married	292.29	1.12	1.66		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.5.2.3: Initial One Year Estimates of Factors That Moderate Head Start's Impact on CTOPPP Elision: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

CTOPPP Elision					
	Intent-To-Treat Impact Estimates				
Moderator/Subgroup (Sample N=1,638)	Non-Head Start Group Mean	Impact of Head Start on Subgroup		Difference in Impact Between Subgroups	
		Demographic Covariate Only	With Fall Measure	Demographic Covariate Only	With Fall Measure
Special Needs				1.78	0.79
Special Needs	237.09	1.95	3.04		
No Special Needs	240.07	3.73	4.19		
Child's Race					
White	243.81	5.81	7.36	1.57 (vs. Black)	0.11 (vs. Black)
Black	240.81	7.39	7.47*	10.49 (vs. Hispanic)	10.38 (vs. Hispanic)
Hispanic	233.88	-3.10	-2.91	8.91 (vs. White)	10.27 (vs. White)
Caregiver Depression (Continuous)				-0.14*	-0.13*
Language of Assessment				6.78	7.22
English-English	242.93	4.50	5.31		
Spanish-English	224.55	-2.28	-1.91		
Parent Married				5.11	4.65
Married	236.22	6.34	6.47		
Not Married	242.67	1.23	1.82		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.5.2.4: Initial One Year Estimates of Factors That Moderate Head Start's Impact on CTOPPP Elision: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

CTOPPP Elision					
	Intent-To-Treat Impact Estimates				
Moderator/Subgroup (Sample N=1,638)	**Non-Head Start Group Mean**	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				4.83	5.94
Special Needs	266.55	-3.26	-3.86		
No Special Needs	274.55	1.57	2.09		
Child's Race					
White	294.13	-3.22	-3.77	8.47 (vs. Black)	9.56 (vs. Black)
Black	275.88	5.25	5.80	3.43 (vs. Hispanic)	2.91 (vs. Hispanic)
Hispanic	256.03	1.82	2.89	5.04 (vs. White)	6.65 (vs. White)
Caregiver Depression (Continuous)				-0.02	-0.02
Language of Assessment				1.99	2.50
English-English	282.77	1.75	1.99		
Spanish-English	251.73	-0.24	-0.51		
Parent Married				1.90	2.13
Married	272.38	2.27	2.63		
Not Married	274.71	0.37	0.50		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.5.2.5: Initial One Year Estimates of Factors That Moderate Head Start's Impact on Woodcock-Johnson Oral Comprehension: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Woodcock-Johnson Oral Comprehension					
	Intent-To-Treat Impact Estimates				
Moderator/Subgroup (Sample N=1,638)	Non-Head Start Group Mean	Impact of Head Start on Subgroup		Difference in Impact Between Subgroups	
		Demographic Covariate Only	With Fall Measure	Demographic Covariate Only	With Fall Measure
Special Needs				0.74	0.59
Special Needs	435.21	-0.31	0.10		
No Special Needs	435.46	0.43	0.69		
Child's Race					
White	440.63	2.22*	2.82**	4.24* (vs. Black)	4.73* (vs. Black)
Black	437.71	-2.02	-1.91	3.03 (vs. Hispanic)	3.05 (vs. Hispanic)
Hispanic	427.12	1.01	1.14	1.21 (vs. White)	1.68 (vs. White)
Caregiver Depression (Continuous)				-0.01	-0.00
Language of Assessment				0.13	0.38
English-English	438.38	0.39	0.71		
Spanish-English	421.31	0.27	0.33		
Parent Married				2.63	2.49
Married	434.13	1.98*	2.09*		
Not Married	436.54	-0.65	-0.40		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.5.2.6: Initial One Year Estimates of Factors That Moderate Head Start's Impact on Woodcock-Johnson Oral Comprehension: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Woodcock-Johnson Oral Comprehension					
Moderator/Subgroup (Sample N=1,638)	Intent-To-Treat Impact Estimates				
	Non-Head Start Group Mean	Impact of Head Start on Subgroup		Difference in Impact Between Subgroups	
		Demographic Covariate Only	With Fall Measure	Demographic Covariate Only	With Fall Measure
Special Needs				2.66	2.85
Special Needs	442.34	-3.40	-3.42		
No Special Needs	443.81	-0.74	-0.56		
Child's Race					
White	454.36	-0.95	-1.38	380.19 (vs. Black)	0.83 (vs. Black)
Black	448.58	-0.76	-0.53	0.57 (vs. Hispanic)	0.21 (vs. Hispanic)
Hispanic	432.06	-1.33	-0.74	0.38 (vs. White)	0.64 (vs. White)
Caregiver Depression (Continuous)				0.01	0.01
Language of Assessment				0.73	1.00
English-English	450.56	-0.68	-0.61		
Spanish-English	426.65	-1.41	-1.61		
Parent Married				1.64	1.83
Married	442.02	-1.90	-1.92		
Not Married	444.98	-0.26	-0.09		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.5.2.7: Initial One Year Estimates of Factors That Moderate Head Start's Impact on Woodcock-Johnson Spelling: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Woodcock-Johnson Spelling					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				0.90	0.68
Special Needs	340.39	2.85	3.00		
No Special Needs	334.05	1.94	2.32		
Child's Race					
White	345.11	0.76	1.49	1.04 (vs. Black)	1.41 (vs. Black)
Black	342.99	-0.28	0.08	5.96 (vs. Hispanic)	5.54 (vs. Hispanic)
Hispanic	342.79	5.68*	5.61*	4.92 (vs. White)	4.12 (vs. White)
Caregiver Depression (Continuous)				-0.04	-0.04
Language of Assessment				5.24	4.92
English-English	344.44	0.95	1.41		
Spanish-English	340.35	6.19	6.33		
Parent Married				1.12	1.11
Married	345.42	2.80	3.09		
Not Married	342.08	1.68	1.98		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.5.2.8: Initial One Year Estimates of Factors That Moderate Head Start's Impact on Woodcock-Johnson Spelling: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Woodcock-Johnson Spelling					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				6.77	7.19
Special Needs	362.28	-1.98	-2.22		
No Special Needs	368.34	4.79*	4.97**		
Child's Race					
White	368.98	2.07	1.83	7.39 (vs. Black)	7.92 (vs. Black)
Black	363.88	9.46**	9.75**	7.25 (vs. Hispanic)	7.21 (vs. Hispanic)
Hispanic	368.73	2.22	2.55	0.15 (vs. White)	0.71 (vs. White)
Caregiver Depression (Continuous)				-0.02	-0.02
Language of Assessment				0.62	1.08
English-English	367.16	4.30*	4.49**		
Spanish-English	368.83	3.68	3.41		
Parent Married				4.80	4.64
Married	371.09	1.53	1.67		
Not Married	364.74	6.33*	6.31*		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.5.2.9: Initial One Year Estimates of Factors That Moderate Head Start's Impact on Woodcock-Johnson Letter-Word Identification: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Woodcock-Johnson Letter-Word Identification					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				1.29	1.25
Special Needs	297.49	6.78	6.63		
No Special Needs	300.90	5.49*	5.38**		
Child's Race					
White	302.11	4.92	3.79	0.30 (vs. Black)	2.00 (vs. Black)
Black	304.71	4.62	5.80**	2.80 (vs. Hispanic)	1.12 (vs. Hispanic)
Hispanic	294.56	7.42**	6.92**	2.50 (vs. White)	3.12 (vs. White)
Caregiver Depression (Continuous)				-0.07	-0.06
Language of Assessment				1.83	2.33
English-English	302.69	5.18**	5.05***		
Spanish-English	291.26	7.01	7.38		
Parent Married				2.94	1.32
Married	300.62	7.77**	6.53**		
Not Married	300.42	4.84*	5.21**		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.5.2.10: Initial One Year Estimates of Factors That Moderate Head Start's Impact on Woodcock-Johnson Letter-Word Identification: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Woodcock-Johnson Letter-Word Identification					
	Intent-To-Treat Impact Estimates				
Moderator/Subgroup (Sample N=1,638)	Non-Head Start Group Mean	Impact of Head Start on Subgroup		Difference in Impact Between Subgroups	
		Demographic Covariate Only	With Fall Measure	Demographic Covariate Only	With Fall Measure
Special Needs				1.29	0.67
Special Needs	310.12	4.59	4.31		
No Special Needs	320.35	5.88*	4.98*		
Child's Race					
White	322.73	5.28	0.26	5.28 (vs. Black)	7.29 (vs. Black)
Black	326.72	10.56*	7.55	7.45 (vs. Hispanic)	0.51 (vs. Hispanic)
Hispanic	312.40	3.11	7.04*	2.17 (vs. White)	6.78 (vs. White)
Caregiver Depression (Continuous)				-0.04	0.00
Language of Assessment				4.76	3.47
English-English	322.49	7.32*	3.92		
Spanish-English	311.73	2.55	7.40*		
Parent Married				4.85	2.59
Married	319.33	3.07	3.59		
Not Married	319.13	7.92*	6.17*		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.5.2.11: Initial One Year Estimates of Factors That Moderate Head Start's Impact on Letter Naming Task: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Letter Naming Task					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				0.58	0.49
Special Needs	3.28	1.73	1.73		
No Special Needs	4.00	1.15*	1.24*		
Child's Race					
White	3.70	1.15	1.33	0.10 (vs. Black)	0.20 (vs. Black)
Black	5.54	1.06	1.13	0.40 (vs. Hispanic)	0.32 (vs. Hispanic)
Hispanic	2.50	1.46*	1.45*	0.30 (vs. White)	0.12 (vs. White)
Caregiver Depression (Continuous)				-0.02	-0.02
Language of Assessment				0.59	0.55
English-English	4.32	1.09	1.19		
Spanish-English	2.21	1.68	1.74		
Parent Married				0.03	-0.00
Married	4.38	1.37	1.46		
Not Married	3.53	1.40*	1.46*		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.5.2.12: Initial One Year Estimates of Factors That Moderate Head Start's Impact on Letter Naming Task: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Letter Naming Task					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				0.73	0.95
Special Needs	6.47	1.59	1.45		
No Special Needs	9.56	2.32*	2.39**		
Child's Race					
White	9.56	2.88**	2.77**	0.31 (vs. Black)	0.57 (vs. Black)
Black	12.12	3.18	3.34	2.06 (vs. Hispanic)	2.10 (vs. Hispanic)
Hispanic	7.34	1.12	1.24	1.75 (vs. White)	1.53 (vs. White)
Caregiver Depression (Continuous)				-0.02	-0.02
Language of Assessment				2.31	2.54
English-English	10.06	2.93**	3.05**		
Spanish-English	7.32	0.62	0.51		
Parent Married				0.72	0.64
Married	9.10	2.00	2.06		
Not Married	9.30	2.72*	2.70*		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.5.2.13: Initial One Year Estimates of Factors That Moderate Head Start's Impact on Woodcock-Johnson Applied Problems: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Woodcock-Johnson Applied Problems					
	Intent-To-Treat Impact Estimates				
Moderator/Subgroup (Sample N=1,638)	**Non-Head Start Group Mean**	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				1.66	1.94
Special Needs	370.54	2.16	2.33		
No Special Needs	373.96	3.82	4.27		
Child's Race					
White	382.40	5.26	6.47	5.99 (vs. Black)	7.17 (vs. Black)
Black	375.37	-0.73	-0.69	7.29 (vs. Hispanic)	7.27 (vs. Hispanic)
Hispanic	362.47	6.56	6.57	1.30 (vs. White)	0.10 (vs. White)
Caregiver Depression (Continuous)				-0.09*	-0.09*
Language of Assessment				4.02	3.55
English-English	378.25	2.81	3.40		
Spanish-English	353.45	6.83	6.96		
Parent Married				2.04	1.74
Married	374.52	5.24	5.39		
Not Married	372.76	3.21	3.65		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.5.2.14: Initial One Year Estimates of Factors That Moderate Head Start's Impact on Woodcock-Johnson Applied Problems: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Woodcock-Johnson Applied Problems					
Moderator/Subgroup (Sample N=1,638)	Intent-To-Treat Impact Estimates				
	Non-Head Start Group Mean	Impact of Head Start on Subgroup		Difference in Impact Between Subgroups	
		Demographic Covariate Only	With Fall Measure	Demographic Covariate Only	With Fall Measure
Special Needs				2.22	2.25
Special Needs	391.46	0.72	0.89		
No Special Needs	394.78	2.94	3.14		
Child's Race					
White	404.36	2.71	2.42	1.27 (vs. Black)	0.72 (vs. Black)
Black	397.11	1.43	1.70	2.01 (vs. Hispanic)	2.29 (vs. Hispanic)
Hispanic	385.32	3.45	3.99	0.74 (vs. White)	1.57 (vs. White)
Caregiver Depression (Continuous)				-0.03	-0.04
Language of Assessment				3.36	2.94
English-English	400.81	1.85	2.06		
Spanish-English	379.87	5.20	5.00		
Parent Married				1.60	1.49
Married	394.23	3.99	4.06		
Not Married	394.59	2.39	2.57		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.5.2.15: Initial One Year Estimates of Factors That Moderate Head Start's Impact on Counting Bears: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Counting Bears					
	Intent-To-Treat Impact Estimates				
Moderator/Subgroup (Sample N=1,638)	Non-Head Start Group Mean	Impact of Head Start on Subgroup		Difference in Impact Between Subgroups	
		Demographic Covariate Only	With Fall Measure	Demographic Covariate Only	With Fall Measure
Special Needs				0.08	0.08
Special Needs	2.52	0.07	0.04		
No Special Needs	2.70	0.15	0.13		
Child's Race					
White	2.91	0.08	-0.02	0.02 (vs. Black)	0.16 (vs. Black)
Black	2.76	0.10	0.14	0.14 (vs. Hispanic)	0.08 (vs. Hispanic)
Hispanic	2.37	0.24	0.22	0.16 (vs. White)	0.24 (vs. White)
Caregiver Depression (Continuous)				-0.00*	-0.00*
Language of Assessment				0.14	0.03
English-English	2.80	0.11	0.12		
Spanish-English	2.20	0.25	0.09		
Fall Bear Rate					0.04
Parent Married				0.08	0.08
Married	2.78	0.11	0.08		
Not Married	2.60	0.19*	0.17		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.5.2.16: Initial One Year Estimates of Factors That Moderate Head Start's Impact on Counting Bears: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Counting Bears					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				0.13	0.15
Special Needs	3.26	0.28	0.25		
No Special Needs	3.64	0.15	0.10		
Child's Race					
White	3.95	-0.14	-0.19	0.41 (vs. Black)	0.39 (vs. Black)
Black	3.59	0.27	0.19	0.06 (vs. Hispanic)	0.13 (vs. Hispanic)
Hispanic	3.34	0.33	0.32	0.47 (vs. White)	0.52* (vs. White)
Caregiver Depression (Continuous)				0.00	0.00
Language of Assessment				0.39	0.24
English-English	3.78	0.04	0.05		
Spanish-English	3.20	0.42	0.28		
Fall Bear Rate					0.03
Parent Married				0.15	0.24
Married	3.65	0.24	0.25		
Not Married	3.56	0.09	0.01		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.5.2.17: Initial One Year Estimates of Factors That Moderate Head Start's Impact on McCarthy Drawing: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

McCarthy Drawing					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				0.03	0.04
Special Needs	2.85	0.11	0.12		
No Special Needs	3.07	0.15	0.16*		
Child's Race					
White	3.00	0.16	0.18	0.03 (vs. Black)	0.00 (vs. Black)
Black	2.96	0.18*	0.18*	0.10 (vs. Hispanic)	0.08 (vs. Hispanic)
Hispanic	3.17	0.09	0.10	0.07 (vs. White)	0.08 (vs. White)
Caregiver Depression (Continuous)				-0.00	-0.00
Language of Assessment				0.04	0.07
English-English	3.04	0.13*	0.17*		
Spanish-English	3.08	0.18	0.10		
Fall Draw Score					-0.02
Parent Married				0.05	0.06
Married	3.06	0.17	0.18		
Not Married	3.03	0.12	0.12		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.5.2.18: Initial One Year Estimates of Factors That Moderate Head Start's Impact on McCarthy Drawing: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

McCarthy Drawing					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				0.18	0.08
Special Needs	3.63	0.37	0.20		
No Special Needs	4.47	0.20	0.12		
Child's Race					
White	4.25	0.28	0.17	0.12 (vs. Black)	0.14 (vs. Black)
Black	4.30	0.16	0.03	0.05 (vs. Hispanic)	0.11 (vs. Hispanic)
Hispanic	4.52	0.20	0.15	0.08 (vs. White)	0.02 (vs. White)
Caregiver Depression (Continuous)				-0.01	-0.00
Language of Assessment				0.09	0.11
English-English	4.26	0.25	0.09		
Spanish-English	4.63	0.17	0.20		
Parent Married				0.35	0.20
Married	4.56	0.04	0.02		
Not Married	4.22	0.39*	0.23		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.5.2.19: Initial One Year Estimates of Factors That Moderate Head Start's Impact on Color Naming/Identification: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Color Naming/Identification					
	Intent-To-Treat Impact Estimates				
Moderator/Subgroup (Sample N=1,638)	Non-Head Start Group Mean	Impact of Head Start on Subgroup		Difference in Impact Between Subgroups	
		Demographic Covariate Only	With Fall Measure	Demographic Covariate Only	With Fall Measure
Special Needs				0.45	0.10
Special Needs	11.91	0.47	0.61		
No Special Needs	13.18	0.92	0.71		
Child's Race					
White	15.24	0.04	0.10	0.47 (vs. Black)	0.77 (vs. Black)
Black	13.37	0.51	0.87	1.52 (vs. Hispanic)	0.22 (vs. Hispanic)
Hispanic	10.45	2.03*	1.09	1.99* (vs. White)	0.99 (vs. White)
Caregiver Depression (Continuous)				-0.02**	-0.02*
Language of Assessment				1.32	0.86
English-English	14.07	0.59	0.87**		
Spanish-English	8.71	1.91	0.01		
Fall Color Score					0.00
Parent Married					
Married	12.76	2.02**	1.50**	1.86*	1.29
Not Married	13.27	0.16	0.22		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.5.2.20: Initial One Year Estimates of Factors That Moderate Head Start's Impact on Color Naming/Identification: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Color Naming/Identification					
Moderator/Subgroup (Sample N=1,638)	Intent-To-Treat Impact Estimates				
	Non-Head Start Group Mean	Impact of Head Start on Subgroup		Difference in Impact Between Subgroups	
		Demographic Covariate Only	With Fall Measure	Demographic Covariate Only	With Fall Measure
Special Needs				0.97	0.98
Special Needs	15.02	1.46	1.05		
No Special Needs	16.63	0.49	0.07		
Child's Race					
White	18.20	0.28	0.03	0.51 (vs. Black)	0.06 (vs. Black)
Black	16.86	0.79	0.08	0.06 (vs. Hispanic)	0.27 (vs. Hispanic)
Hispanic	14.90	0.73	0.36	0.45 (vs. White)	0.33 (vs. White)
Caregiver Depression (Continuous)				-0.00	-0.00
Language of Assessment				0.40	0.32
English-English	17.48	0.50	0.08		
Spanish-English	14.15	0.90	0.40		
Parent Married				0.31	0.37
Married	16.42	0.87	0.48		
Not Married	16.48	0.57	0.12		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.5.2.21: Initial One Year Estimates of Factors That Moderate Head Start's Impact on PELS: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

PELS					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				0.24	0.11
Special Needs	2.22	0.26	0.32		
No Special Needs	2.37	0.50***	0.43***		
Child's Race					
White	2.50	0.37***	0.32**	0.16 (vs. Black)	0.25 (vs. Black)
Black	2.48	0.53**	0.57***	0.02 (vs. Hispanic)	0.23 (vs. Hispanic)
Hispanic	2.08	0.51**	0.35*	0.14 (vs. White)	0.02 (vs. White)
Caregiver Depression (Continuous)				-0.00*	-0.00*
Language of Assessment				0.02	0.12
English-English	2.47	0.48***	0.44***		
Spanish-English	1.89	0.46*	0.32		
Parent Married				0.09	0.01
Married	2.35	0.52***	0.43***		
Not Married	2.36	0.43***	0.41***		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.5.2.22: Initial One Year Estimates of Factors That Moderate Head Start's Impact on PELS: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

PELS					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				0.15	0.03
Special Needs	3.12	0.28	0.32		
No Special Needs	3.35	0.43***	0.28**		
Child's Race					
White	3.70	0.23	0.07	0.51 (vs. Black)	0.36 (vs. Black)
Black	3.40	0.75**	0.44	0.39 (vs. Hispanic)	0.07 (vs. Hispanic)
Hispanic	3.00	0.35	0.36	0.12 (vs. White)	0.29 (vs. White)
Caregiver Depression (Continuous)				-0.00	0.00
Language of Assessment				0.12	0.13
English-English	3.50	0.45***	0.25*		
Spanish-English	2.95	0.33	0.38		
Parent Married				0.16	0.03
Married	3.31	0.35*	0.27*		
Not Married	3.34	0.52**	0.30*		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Appendix 6.1: Social-Emotional Domain Estimated Impact on Program Participants

This appendix focuses on the impact estimates that are adjusted for the fact that some of the children assigned to Head Start failed to take advantage of this opportunity; i.e., they never participated in Head Start (the "no-shows") and that some children assigned to the non-Head Start group managed to find their way into the program (the "crossovers"). These results are summarized in Exhibit A.6.1.1 and A.6.1.2.

All measures of the impact of **participation** in Head Start exceeded the corresponding measures of the average impact of **access to** Head Start, regardless of which adjustment method is used. For example, the first row of Exhibit A.6.1.1 shows an estimated impact of -0.48 on the Total Problem Behavior Scale for the average child in the 3-year-old group granted access to the program and a larger average impact of -0.54 to -0.60 for those children who actually attended Head Start. These same increases are again seen in the relationship between the calculated effect sizes for the estimates of "access" vs. "participation," with a few instances of the size of the effect moving from "small" (under 0.2) to "modest" (between 0.2 and 0.5).

A comparison of the estimates resulting from the two adjustment methods indicates that there is, as expected, some dampening effect of the crossover children in the non-Head Start group, i.e., reducing the size of any observed difference in outcomes between the Head Start and non-Head Start groups. However, as previously noted, the combined no-show/crossover adjustments should only be used as a rough indication of the likely consequences of the presence of crossovers on the interpretation of the impact of Head Start on children's school readiness.

Exhibit A.6.1.1: Initial Estimates of the Impact of Head Start on Social-Emotional Outcomes, Intent to Treat, and Impact on the Treated: Statistically Significant Results Only, 3-Year-Old Group, Combined English-English and Spanish-English Group (Weighted Data)

Outcome Measure	Impact Estimates			Effect Sizes		
	Impact of Access to Head Start	**Impact of Head Start Participation, No-Show Adjustment**	**Impact of Head Start Participation, Combined No-Show and Crossover Adjustment**	**Impact of Access to Head Start**	**Impact of Head Start Participation, No-Show Adjustment**	**Impact of Head Start Participation, Combined No-Show and Crossover Adjustment**
Overall Impact						
Total Problem Behavior Scale	-0.48**	-0.54**	-0.60**	-0.13	-0.14	-0.16
Hyperactive Behavior Scale	-0.29**	-0.33**	-0.32*	- 0.18	- 0.21	-0.20
Difference in Impact [1]						
Social Competencies Checklist: Race (White Impact Exceeds African American)	0.49**	0.56***	N/A	0.39	0.45	N/A
Social Competencies Checklist: Race (Hispanic Impact Exceeds African American)	0.37*	0.42*	N/A	0.30	0.34	N/A
Impact on Subgroup [2]						
Total Problem Behavior Scale: No Special Needs	-0.52*	-0.59*	N/A	-0.14	-0.16	N/A
Total Problem Behavior Scale: White	-0.86**	-0.98*	N/A	-0.23	-0.26	N/A
Total Problem Behavior Scale: Parents Not Separated or Divorced	-0.50**	-0.56**	N/A	-0.13	-0.15	N/A
Total Problem Behavior Scale: Parent Not Married	-0.47*	-0.52*	N/A	-0.13	-0.14	N/A
Total Problem Behavior Scale: English-English Language Group	-0.46*	-0.52*	N/A	-0.12	-0.14	N/A
Aggressive Behavior: White	-0.30*	-0.34*	N/A	-0.17	-0.19	N/A
Hyperactive Behavior Scale: No Special Needs	-0.30*	-0.34*	N/A	-0.19	-0.22	N/A
Hyperactive Behavior Scale: White	-0.34*	-0.39*	N/A	-0.22	-0.25	N/A

Exhibit A.6.1.1: (continued)

Outcome Measure	Impact Estimates			Effect Sizes		
	Impact of Access to Head Start	Impact of Head Start Participation, No-Show Adjustment	Impact of Head Start Participation, Combined No-Show and Crossover Adjustment	Impact of Access to Head Start	Impact of Head Start Participation, No-Show Adjustment	Impact of Head Start Participation, Combined No-Show and Crossover Adjustment
Hyperactive Behavior Scale: Hispanic	-0.40*	-0.45**	N/A	-0.25	-0.28	N/A
Hyperactive Behavior Scale: Male	-0.31*	-0.36*	N/A	-0.19	-0.23	N/A
Hyperactive Behavior Scale: Parent Not Separated or Divorced	-0.33***	-0.37***	N/A	-0.21	-0.23	N/A
Hyperactive Behavior Scale: Parent Married	-0.39*	-0.45*	N/A	-0.25	-0.28	N/A
Hyperactive Behavior Scale: Parent Not Married	-0.25*	-0.28*	N/A	-0.16	-0.18	N/A
Hyperactive Behavior Scale: English-English Language Group	-0.20*	-0.23*	N/A	-0.13	-0.15	N/A
Hyperactive Behavior Scale: Spanish-English Language Group	-0.68**	-0.77*	N/A	-0.43	-0.49	N/A
Social Competencies Checklist: African American	-0.34**	-0.38**	N/A	-0.27	-0.30	N/A

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.

[1] A total of 60 differences in impacts between subgroups were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 6.2. Findings for depression indicate the change in Head Start's estimated impact that accompanies a 1-point increase in mother's baseline depression score. Findings for baseline factors other than depression indicate the amount by which Head Start's estimated impact for the first subset of participants listed in the row label exceeds that for the second subset listed.

[2] A total of 78 subgroup impacts were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 6.2.

Exhibit A.6.1.2: Initial Estimates of the Impact of Head Start on Social Emotional, Intent to Treat and Impact on the Treated: Statistically Significant Results Only, 4-Year-Old Group, Combined English-English Spanish-English Group (Weighted Data)

Outcome Measure	*Impact Estimates*			*Effect Sizes*		
	Impact of Access to Head Start	*Impact of Head Start Participation, No Show Adjustment*	*Impact of Head Start Participation, Combined No Show and Crossover Adjustment*	*Impact of Access to Head Start*	*Impact of Head Start Participation, No Show Adjustment*	*Impact of Head Start Participation, Combined No Show and Crossover Adjustment*
Overall Impact						
No statistically significant impacts	N/A	N/A	N/A	N/A	N/A	N/A
Difference in Impact [1]						
Social Competencies Checklist: Depression	-0.00*	-0.00*	N/A	- 0.00	- 0.00	N/A
Aggressive Behavior: (African American Impact Exceeds Hispanic)	0.81**	0.97**	N/A	0.51	0.61	N/A
Impact on Subgroup [2]						
Total Problem Behavior Scale: African American	-0.92**	-1.20**	N/A	-0.27	-0.36	N/A
Aggressive Behavior Scale: African American	-0.61**	-0.80**	N/A	-0.38	-0.50	N/A
Aggressive Behavior Scale: Female	-0.30*	-0.37*	N/A	-0.19	-0.23	N/A
Aggressive Behavior Scale: English-English Language Group	-0.24*	-0.30*	N/A	-0.15	-0.19	N/A

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.

[1] A total of 60 differences in impacts between subgroups were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 6.2. Findings for depression indicate the change in Head Start's estimated impact that accompanies a 1-point increase in mother's baseline depression score. Findings for baseline factors other than depression indicate the amount by which Head Start's estimated impact for the first subset of participants listed in the row label exceeds that for the second subset listed.

[2] A total of 78 subgroup impacts were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 6.2.

Appendix 6.2: Factors That Moderate the Impact of Head Start: Detailed Tables for Social-Emotional Outcomes

The following tables (Exhibits A.6.2.1 through A.6.2.12) provide the results of the moderator/subgroup analyses for measures of social-emotional outcomes, with a separate table for each individual measure. For clarity, these results are only presented for the full combined sample (i.e., not separately for the English-English and Spanish-English language groups). Each table is organized as follows:

- The first column lists the variable used in the particular subgroup/moderator analysis (separate regressions were estimated for each moderator). For example, analyses were conducted to examine the extent to which child gender was related to program impact.
- As shown, separate lines are shown for the overall construct (e.g., gender) and for each of the subgroups that make up the construct, e.g., boys and girls. Estimates associated with the overall construct represent estimated **differences** in impacts (e.g., boys vs. girls), while the figures associated with each of the subgroup rows represent the **impact on the individual subgroups** (e.g., impacts on boys alone).
- For comparison purposes, the next column provides the mean on the particular outcome measure for the group indicated among children in the non-Head Start group in Spring 2003 (the end of the first program year).
- The next set of columns provides the estimated impact on the individual subgroups, while the last two columns provide the estimated difference in impact between subgroups.
- As with the overall impact tables provided in Chapter 6, the estimated impacts are shown using two separate estimation specifications: (1) using regression analyses that include only demographic covariates measured in fall 2002 and (2) using regression analyses that added a measure of the outcome variable assessed in fall 2002. The highlighting indicates which estimate is considered the "best" (see Chapter 4) and which is highlighted in the discussion in Chapter 6.

Exhibit A.6.2.1: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Aggressive Behavior: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Aggressive Behavior					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				0.24	0.07
Special Needs	3.75	-0.28	-0.17		
No Special Needs	2.96	-0.04	-0.11		
Child's Race					
White	3.14	-0.22	-0.30*	0.20 (vs. Black)	0.32 (vs. Black)
Black	2.87	-0.02	0.02	0.04 (vs. Hispanic)	0.10 (vs. Hispanic)
Hispanic	3.15	0.02	-0.08	0.24 (vs. White)	0.21 (vs. White)
Child's Gender				0.16	0.12
Girl	3.03	-0.15	-0.17		
Boy	3.07	0.01	-0.06		
Caregiver Depression (Continuous)				0.00	0.00
Parent Separated or Divorced				0.26	0.33
Separated or Divorced	3.31	-0.29	-0.41		
Not Separated or Divorced	3.01	-0.03	-0.07		
Parent Married				0.17	0.07
Married	2.98	-0.15	-0.15		
Not Married	3.11	0.01	-0.08		
PPVT (Continuous)					0.00
Language of Assessment				0.17	0.14
English-English	2.99	-0.11	-0.15		
Spanish-English	3.32	0.06	-0.01		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.6.2.2: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Aggressive Behavior: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Aggressive Behavior					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				0.34	0.36
Special Needs	3.17	0.16	0.28		
No Special Needs	2.83	-0.18	-0.08		
Child's Race					
White	2.88	-0.20	-0.01	0.41 (vs. Black)	0.39 (vs. Black)
Black	2.80	-0.61**	-0.40*	0.81** (vs. Hispanic)	0.55 (vs. Hispanic)
Hispanic	2.90	0.20	0.16	0.40 (vs. White)	0.16 (vs. White)
Child's Gender				0.31	0.26
Girl	2.79	-0.30*	-0.17		
Boy	2.94	0.01	0.09		
Caregiver Depression (Continuous)				0.00	0.00
Parent Separated or Divorced				0.28	0.34
Separated or Divorced	2.90	0.11	0.28		
Not Separated or Divorced	2.87	-0.17	-0.06		
Parent Married				0.08	0.07
Married	2.73	-0.09	0.03		
Not Married	2.99	-0.17	-0.04		
PPVT (Continuous)					0.00
Language of Assessment				0.33	0.10
English-English	2.77	-0.24*	-0.07		
Spanish-English	3.1	0.09	0.03		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.6.2.3: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Social Skills and Positive Approaches to Learning: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Social Skills and Positive Approaches to Learning					
	Intent-To-Treat Impact Estimates				
Moderator/Subgroup (Sample N=1,638)	**Non-Head Start Group Mean**	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				0.12	0.00
Special Needs	11.90	0.17	-0.00		
No Special Needs	12.43	0.04	-0.00		
Child's Race					
White	12.22	0.37*	0.22	0.47* (vs. Black)	0.36 (vs. Black)
Black	12.33	-0.09	-0.14	0.00 (vs. Hispanic)	0.05 (vs. Hispanic)
Hispanic	12.56	-0.09	-0.09	0.46* (vs. White)	0.31 (vs. White)
Child's Gender				0.09	0.01
Girl	12.58	0.01	0.00		
Boy	12.15	0.11	-0.01		
Caregiver Depression (Continuous)				0.00	-0.00
Parent Separated or Divorced				0.57*	0.38
Separated or Divorced	12.74	-0.43	-0.30		
Not Separated or Divorced	12.31	0.14	0.08		
Parent Married				0.06	0.07
Married	12.35	0.10	-0.01		
Not Married	12.39	0.05	0.06		
PPVT (Continuous)					0.00
Language of Assessment				0.38	0.31
English-English	12.33	0.13	0.06		
Spanish-English	12.54	-0.25	-0.25		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.6.2.4: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Social Skills and Positive Approaches to Learning: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Social Skills and Positive Approaches to Learning					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				0.09	0.28
Special needs	12.10	0.07	0.19		
No special needs	12.54	-0.01	-0.09		
Child's Race					
White	12.58	0.08	0.00	0.13 (vs. Black)	0.09 (vs. Black)
Black	12.39	0.22	0.09	0.43 (vs. Hispanic)	0.28 (vs. Hispanic)
Hispanic	12.47	-0.22	-0.19	0.30 (vs. White)	0.19 (vs. White)
Child's Gender				0.30	0.30
Girl	12.54	0.15	0.10		
Boy	12.44	-0.15	-0.20		
Caregiver Depression (Continuous)				0.00	-0.00
Parent Separated or Divorced				0.02	0.07
Separated or Divorced	12.39	-0.00	0.03		
Not Separated or Divorced	12.51	0.02	-0.04		
Parent Married				0.22	0.22
Married	12.38	0.13	0.10		
Not Married	12.58	-0.09	-0.13		
PPVT (Continuous)					0.00
Language of Assessment				0.46	0.39
English-English	12.47	0.13	0.06		
Spanish-English	12.52	-0.33	-0.33		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.6.2.5: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Total Child Behavior Problems: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Total Child Behavior Problems					
	Intent-To-Treat Impact Estimates				
Moderator/Subgroup (Sample N=1,638)	**Non-Head Start Group Mean**	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				0.08	0.33
Special Needs	7.85	-0.56	-0.19		
No Special Needs	6.06	-0.48*	-0.52*		
Child's Race					
White	6.57	-1.00**	-0.86**	0.79 (vs. Black)	0.58 (vs. Black)
Black	5.57	-0.21	-0.28	0.08 (vs. Hispanic)	0.04 (vs. Hispanic)
Hispanic	6.63	-0.29	-0.33	0.72 (vs. White)	0.53 (vs. White)
Child's Gender				0.10	0.01
Girl	5.91	-0.45	-0.49		
Boy	6.62	-0.54	-0.48		
Caregiver Depression (Continuous)				0.00	-0.00
Parent Separated or Divorced				0.13	0.27
Separated or Divorced	6.53	-0.66	-0.77		
Not Separated or Divorced	6.22	-0.53*	-0.50**		
Parent Married				0.57	0.15
Married	6.2	-0.86*	-0.62		
Not Married	6.3	-0.29	-0.47*		
PPVT (Continuous)					-0.00
Language of Assessment				0.15	0.16
English-English	5.99	-0.53**	-0.46*		
Spanish-English	7.36	-0.38	-0.62		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.6.2.6: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Total Child Behavior Problems: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Total Child Behavior Problems					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				0.58	0.58
Special Needs	7.60	0.26	0.50		
No Special Needs	5.61	-0.32	-0.08		
Child's Race					
White	5.61	-0.20	0.25	0.72 (vs. Black)	0.73 (vs. Black)
Black	5.35	-0.92**	0.48	1.04 (vs. Hispanic)	0.55 (vs. Hispanic)
Hispanic	6.28	0.12	0.07	0.32 (vs. White)	0.18 (vs. White)
Child's Gender				0.53	0.60
Girl	5.42	-0.52	-0.32		
Boy	6.22	0.01	0.29		
Caregiver Depression (Continuous)				0.01	0.00
Parent Separated or Divorced				0.19	0.27
Separated or Divorced	6.2	-0.44	-0.23		
Not Separated or Divorced	5.77	-0.25	0.04		
Parent Married				0.41	0.65
Married	5.56	0.05	0.35		
Not Married	6.07	-0.46	-0.30		
PPVT (Continuous)					0.00
Language of Assessment				0.54	0.17
English-English	5.46	-0.41	-0.06		
Spanish-English	6.66	0.13	0.11		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.6.2.7: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Hyperactive Behavior: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Hyperactive Behavior					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				0.05	0.03
Special Needs	2.59	-0.39	-0.27		
No Special Needs	1.94	-0.33**	-0.30*		
Child's Race					
White	2.01	-0.44**	-0.34*	0.28 (vs. Black)	0.19 (vs. Black)
Black	1.69	-0.16	-0.15	0.27 (vs. Hispanic)	0.25 (vs. Hispanic)
Hispanic	2.34	-0.43**	-0.40*	0.01 (vs. White)	0.06 (vs. White)
Child's Gender				0.09	0.04
Girl	1.87	-0.29	-0.28		
Boy	2.17	-0.39**	-0.31*		
Caregiver Depression (Continuous)				-0.00	-0.00
Parent Separated or Divorced				0.15	0.17
Separated or Divorced	2.09	-0.24	-0.17		
Not Separated or Divorced	2.00	-0.39***	-0.33***		
Parent Married				0.23	0.14
Married	2.00	-0.50**	-0.39*		
Not Married	2.02	-0.26*	-0.25*		
PPVT (Continuous)					-0.00
Language of Assessment				0.30	0.47
English-English	1.83	-0.28**	-0.20*		
Spanish-English	2.79	-0.59**	-0.68**		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.6.2.8: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Hyperactive Behavior: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Hyperactive Behavior					
	Intent-To-Treat Impact Estimates				
		Impact of Head Start on Subgroup		Difference in Impact Between Subgroups	
Moderator/Subgroup (Sample N=1,638)	**Non-Head Start Group Mean**	**Demographic Covariate Only**	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				0.23	0.33
Special Needs	2.38	0.12	0.29		
No Special Needs	1.71	-0.11	-0.04		
Child's Race					
White	1.67	-0.11	0.09	0.03 (vs. Black)	0.16 (vs. Black)
Black	1.47	-0.15	-0.07	0.12 (vs. Hispanic)	0.05 (vs. Hispanic)
Hispanic	2.06	-0.03	-0.03	0.09 (vs. White)	0.12 (vs. White)
Child's Gender				0.11	0.22
Girl	1.58	-0.14	-0.11		
Boy	1.98	-0.03	0.11		
Caregiver Depression (Continuous)				0.00	0.00
Parent Separated or Divorced				0.47	0.44
Separated or Divorced	1.97	-0.51	-0.39		
Not Separated or Divorced	1.76	-0.04	0.05		
Parent Married				0.24	0.33
Married	1.71	0.02	0.16		
Not Married	1.85	-0.22	-0.16		
PPVT (Continuous)					0.00
Language of Assessment				0.16	0.03
English-English	1.57	-0.13	-0.01		
Spanish-English	2.26	0.03	0.02		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.6.2.9: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Social Competencies Checklist: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Social Competencies Checklist					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				0.19	0.23
Special Needs	10.81	-0.20	-0.26		
No Special Needs	11.02	-0.01	-0.03		
Child's Race					
White	10.83	0.20	0.15	0.53** (vs. Black)	0.49** (vs. Black)
Black	11.02	-0.33*	-0.34**	0.36 (vs. Hispanic)	0.37* (vs. Hispanic)
Hispanic	11.14	0.04	0.03	0.16 (vs. White)	0.12 (vs. White)
Child's Gender				0.05	0.06
Girl	11.10	-0.01	-0.03		
Boy	10.94	-0.06	-0.09		
Caregiver Depression (Continuous)				-0.00	-0.00
Parent Separated or Divorced				0.04	0.12
Separated or Divorced	10.96	0.01	0.07		
Not Separated or Divorced	11.00	-0.03	-0.05		
Parent Married				0.10	0.02
Married	11.04	0.03	-0.05		
Not Married	10.96	-0.07	-0.03		
PPVT (Continuous)					-0.00
Language of Assessment				0.06	0.11
English-English	10.94	-0.05	-0.08		
Spanish-English	11.22	0.01	0.03		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.6.2.10: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Social Competencies Checklist: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Social Competencies Checklist					
	Intent-To-Treat Impact Estimates				
Moderator/Subgroup (Sample N=1,638)	Non-Head Start Group Mean	Impact of Head Start on Subgroup		Difference in Impact Between Subgroups	
		Demographic Covariate Only	With Fall Measure	Demographic Covariate Only	With Fall Measure
Special Needs				0.33	0.26
Special Needs	10.84	-0.32	-0.25		
No Special Needs	11.09	0.01	0.01		
Child's Race					
White	11.06	0.10	0.10	0.20 (vs. Black)	0.28 (vs. Black)
Black	11.05	-0.10	-0.18	0.01 (vs. Hispanic)	0.16 (vs. Hispanic)
Hispanic	11.07	-0.09	-0.02	0.19 (vs. White)	0.12 (vs. White)
Child's Gender				0.34	0.33
Girl	11.19	0.15	0.15		
Boy	10.94	-0.19	-0.18		
Caregiver Depression (Continuous)				-0.00*	-0.00**
Parent Separated or Divorced				0.08	0.12
Separated or Divorced	10.96	0.04	0.10		
Not Separated or Divorced	11.1	-0.04	-0.02		
Parent Married				0.04	0.05
Married	11.17	-0.00	-0.03		
Not Married	10.97	-0.05	0.02		
PPVT (Continuous)					0.00
Language of Assessment				0.07	0.01
English-English	11.04	-0.01	-0.02		
Spanish-English	11.11	-0.08	-0.03		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.6.2.11: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Withdrawn Behavior Scale: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Withdrawn Behavior Scale					
Moderator/Subgroup (Sample N=1638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				0.06	0.08
Special Needs	0.76	0.02	0.04		
No Special Needs	0.56	-0.04	-0.04		
Child's Race					
White	0.72	-0.13	-0.08	0.10 (vs. Black)	0.02 (vs. Black)
Black	0.44	-0.03	-0.09	0.10 (vs. Hispanic)	0.16 (vs. Hispanic)
Hispanic	0.57	0.07	0.07	0.20 (vs. White)	0.15 (vs. White)
Child's Gender				0.01	0.02
Girl	0.53	-0.03	-0.02		
Boy	0.63	-0.04	-0.04		
Caregiver Depression (Continuous)				-0.00	-0.00
Parent Separated or Divorced				0.13	0.14
Separated or Divorced	0.65	-0.16	-0.17		
Not Separated or Divorced	0.57	-0.03	-0.03		
Parent Married				0.00	0.07
Married	0.58	-0.05	-0.00		
Not Married	0.57	-0.04	-0.08		
PPVT (Continuous)					-0.00
Language of Assessment				0.36*	0.31
English-English	0.58	-0.10	-0.09		
Spanish-English	0.56	0.26	0.21		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.6.2.12: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Withdrawn Behavior Scale: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Withdrawn Behavior Scale					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				0.10	0.06
Special Needs	1.11	0.03	0.02		
No Special Needs	0.64	-0.07	-0.04		
Child's Race					
White	0.69	-0.01	-0.02	0.10 (vs. Black)	0.03 (vs. Black)
Black	0.62	-0.11	-0.05	0.05 (vs. Hispanic)	0.01 (vs. Hispanic)
Hispanic	0.75	-0.06	-0.04	0.05 (vs. White)	0.03 (vs. White)
Child's Gender				0.10	0.14
Girl	0.67	-0.11	-0.10		
Boy	0.72	-0.00	0.03		
Caregiver Depression (Continuous)				0.00	0.00
Parent Separated or Divorced				0.05	0.12
Separated or Divorced	0.84	-0.10	-0.14		
Not Separated or Divorced	0.67	-0.05	-0.02		
Parent Married				0.01	0.06
Married	0.68	-0.06	-0.00		
Not Married	0.72	-0.07	-0.07		
PPVT (Continuous)					0.00
Language of Assessment				0.09	0.11
English-English	0.68	-0.08	-0.07		
Spanish-English	0.74	0.01	0.04		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Appendix 7.1: Health Domain, Estimated Impact of Program Participation

As in previous chapters, Exhibits A.7.1.1 and A.7.1.2 provide impact estimates that are adjusted for the fact that some of the children assigned to Head Start failed to take advantage of this opportunity—i.e., they never participated in Head Start (the "no shows")─and that some children assigned to the non-Head Start group managed to find their way into the program (the "crossovers"). All measures of the impact of **participation** in Head Start exceeded the corresponding measures of the average impact of **access to** Head Start, regardless of which adjustment method was used. For example, the first row of Exhibit A.7.1.1 shows an estimated impact of 0.17 on whether the child received dental care for the average child in the 3-year-old group granted access to the program through assignment to the Head Start research sample, and a larger average impact of 0.19-0.26 for those children who actually attended Head Start. These same increases are again seen in the relationship between the calculated effect sizes for the estimates of "access" vs. "participation," with a few instances of the size of the effect moving from "modest" (between 0.2 and 0.5) to "large" (over 0.5).

These data provide an indication of the dampening effect of the crossover children in the non-Head Start group, i.e., reducing the size of any observed difference in outcomes between the Head Start and non-Head Start groups. But, again, the combined no-show/crossover adjustments should only be used as a rough indication of the likely consequences of the presence of crossovers on the interpretation of the impact of Head Start on children's school readiness.

Exhibit A.7.1.1: Initial Estimates of the Impact of Head Start on Health Outcomes, Intent to Treat, and Impact on the Treated: Statistically Significant Results Only, 3-Year-Old Group, Combined English-English and Spanish-English Group (Weighted Data)

Outcome Measure	Impact Estimates			Effect Sizes		
	Impact of Access to Head Start	**Impact of Head Start Participation, No-Show Adjustment**	**Impact of Head Start Participation, Combined No-Show and Crossover Adjustment**	**Impact of Access to Head Start**	**Impact of Head Start Participation, No-Show Adjustment**	**Impact of Head Start Participation, Combined No-Show and Crossover Adjustment**
Overall Impact						
Child Health Status Excellent or Very Good	0.05*	0.06*	0.06*	0.12	0.14	0.14
Child Had Dental Care	0.17***	0.19***	0.26***	0.34	0.38	0.52
Difference in Impact [1]						
Child Health Status: Home Language (Non-English Impact Exceeds English)	0.12*	0.14*	N/A	0.28	0.33	N/A
Child Health Status: Depression	0.05*	0.06*	N/A	0.12	0.14	N/A
Child Had Care for Injury: Race (White Impact Exceeds African American)	0.08*	0.09*	N/A	0.30	0.33	N/A
Child Had Care for Injury: Race (White Impact Exceeds Hispanic)	0.13***	0.15***	N/A	0.48	0.56	N/A
Child Had Dental Care: Depression	0.16***	0.18***	N/A	0.32	0.36	N/A
Impact on Subgroup [2]						
Child Health Status: Special Needs	0.19*	0.21*	N/A	0.44	0.49	N/A
Child Health Status: Parent Married	0.08*	0.09*	N/A	0.19	0.21	N/A
Child Health Status: Hispanic	0.12**	0.14**	N/A	0.28	0.33	N/A
Child Health Status: Home Language Not English	0.14**	0.16**	N/A	0.33	0.37	N/A
Child Had Care for Injury: White	0.07***	0.08***	N/A	0.26	0.30	N/A
Child Had Care for Injury: Hispanic	-0.06*	-0.07*	N/A	- 0.22	- 0.26	N/A

Exhibit A.7.1.1: (continued)

Outcome Measure	Impact Estimates			Effect Sizes		
	Impact of Access to Head Start	**Impact of Head Start Participation, No-Show Adjustment**	**Impact of Head Start Participation, Combined No-Show and Crossover Adjustment**	**Impact of Access to Head Start**	**Impact of Head Start Participation, No-Show Adjustment**	**Impact of Head Start Participation, Combined No-Show and Crossover Adjustment**
Child Had Dental Care: Special Needs	0.24*	0.26*	N/A	0.48	0.52	N/A
Child Had Dental Care: No Special Needs	0.16***	0.18***	N/A	0.32	0.36	N/A
Child Had Dental Care: Parent Married	0.18***	0.21***	N/A	0.36	0.42	N/A
Child Had Dental Care: Parent Not Married	0.16***	0.18***	N/A	0.32	0.36	N/A
Child Had Dental Care: White	0.17***	0.19***	N/A	0.34	0.38	N/A
Child Had Dental Care: Hispanic	0.22***	0.25***	N/A	0.44	0.50	N/A
Child Had Dental Care: Home Language Not English	0.22***	0.25***	N/A	0.44	0.50	N/A
Child Had Dental Care: Home Language English	0.15***	0.17***	N/A	0.30	0.34	N/A

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.

[1] A total of 35 differences in impacts between subgroups were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 7.2. Findings for depression indicate the change in Head Start's estimated impact that accompanies a 1-point increase in mother's baseline depression score. Findings for baseline factors other than depression indicate the amount by which Head Start's estimated impact for the first subset of participants listed in the row label exceeds that for the second subset listed.

[2] A total of 50 subgroup impacts were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 7.2.

Exhibit A.7.1.2: Initial Estimates of the Impact of Head Start on Health Outcomes, Intent to Treat, and Impact on the Treated: Statistically Significant Results Only, 4-Year-Old Group, Combined English-English Spanish-English Group (Weighted Data)

Outcome Measure	Impact Estimates			Effect Sizes		
	Impact of Access to Head Start	Impact of Head Start Participation, No-Show Adjustment	Impact of Head Start Participation, Combined No-Show and Crossover Adjustment	Impact of Access to Head Start	Impact of Head Start Participation, No-Show Adjustment	Impact of Head Start Participation, Combined No-Show and Crossover Adjustment
Overall Impact						
Child Had Dental Care	0.16***	0.19***	0.23***	0.32	0.38	0.46
Difference in Impact [1]						
Child Had Health Insurance: Race (Hispanic Impact Exceeds African American)	0.08*	0.10*	N/A	0.24	0.30	N/A
Child Health Status: Special Needs (No Special Needs Impact Exceeds Special Needs)	0.22*	0.26*	N/A	0.56	0.67	N/A
Child Had Dental Care: Depression	0.16***	0.19***	N/A	0.32	0.38	N/A
Impact on Subgroup [2]						
Child Had Health Insurance: Home Language Not English	0.06*	0.07*	N/A	0.18	0.21	N/A
Child Health Status: Special Needs	-0.23*	-0.26*	N/A	-0.59	-0.67	N/A
Child Health Status: Parent Married	-0.08**	-0.10**	N/A	-0.21	-0.26	N/A
Child Health Status: Home Language Not English	-0.08*	-0.09*	N/A	-0.21	-0.23	N/A
Child Had Dental Care: No Special Needs	0.16***	0.19***	N/A	0.32	0.38	N/A

Exhibit A.7.1.2: (continued)

Outcome Measure	Impact Estimates			Effect Sizes		
	Impact of Access to Head Start	**Impact of Head Start Participation, No-Show Adjustment**	**Impact of Head Start Participation, Combined No-Show and Crossover Adjustment**	**Impact of Access to Head Start**	**Impact of Head Start Participation, No-Show Adjustment**	**Impact of Head Start Participation, Combined No-Show and Crossover Adjustment**
Child Had Dental Care: Parent Married	0.18***	0.22***	N/A	0.36	0.44	N/A
Child Had Dental Care: Parent Not Married	0.14**	0.17**	N/A	0.28	0.34	N/A
Child Had Dental Care: White	0.24***	0.28***	N/A	0.48	0.56	N/A
Child Had Dental Care: Hispanic	0.12*	0.14*	N/A	0.24	0.28	N/A
Child Had Dental Care: Home Language Not English	0.17**	0.19**	N/A	0.34	0.38	N/A
Child Had Dental Care: Home Language English	0.16***	0.20***	N/A	0.32	0.40	N/A

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.

[1] A total of 35 differences in impacts between subgroups were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 7.2. Findings for depression indicate the change in Head Start's estimated impact that accompanies a 1-point increase in mother's baseline depression score. Findings for baseline factors other than depression indicate the amount by which Head Start's estimated impact for the first subset of participants listed in the row label exceeds that for the second subset listed.

[2] A total of 50 subgroup impacts were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 7.2.

Appendix 7.2: Factors That Moderate the Impact of Head Start: Detailed Tables for Health Outcomes

The following tables (Exhibits A.7.2.1 through A.7.2.10) provide the results of the moderator/subgroup analyses for measures of health outcomes, with a separate table for each individual measure. For clarity, these results are only presented for the full combined sample (i.e., not separately for the English-English and Spanish-English language groups). Each table is organized as follows:

- The first column lists the variable used in the particular subgroup/moderator analysis (separate regressions were estimated for each moderator). For example, analyses were conducted to examine the extent to which child gender was related to program impact.
- As shown, separate lines are shown for the overall construct (e.g., gender) and for each of the subgroups that make up the construct, e.g., boys and girls. Estimates associated with the overall construct represent estimated **differences** in impacts (e.g., boys vs. girls), while the figures associated with each of the subgroup rows represent the **impact on the individual subgroups** (e.g., impacts on boys alone).
- For comparison purposes, the next column provides the mean on the particular outcome measure for the group indicated among children in the non-Head Start group in spring 2003 (the end of the first program year).
- The next set of columns provides the estimated impact on the individual subgroups, while the last two columns provide the estimated difference in impact between subgroups.
- As with the overall impact tables provided in Chapter 7, the estimated impacts are shown using two separate estimation specifications: (1) using regression analyses that include only demographic covariates measured in fall 2002 and (2) using regression analyses that added a measure of the outcome variable assessed in fall 2002. The highlighting indicates which estimate is considered the "best" (see Chapter 4) and which is highlighted in the discussion in Chapter 7.

Exhibit A.7.2.1: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Health Insurance: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Health Insurance					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				0.03	0.04
Special Needs	0.90	0.03	0.03		
No Special Needs	0.02	-0.00	0.01		
Parent Married				0.00	0.01
Married	0.91	0.00	0.01		
Not Married	0.92	0.00	0.00		
Child's Race					
White	0.94	-0.03	-0.03	0.03 (vs. Black)	0.02 (vs. Black)
Black	0.96	-0.00	-0.01	0.04 (vs. Hispanic)	0.02 (vs. Hispanic)
Hispanic	0.84	0.03	0.02	0.06 (vs. White)	0.04 (vs. White)
Home Language				0.02	0.02
Not English	0.85	0.02	0.01		
English	0.94	-0.01	-0.01		
Depression				0.00	0.00

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.7.2.2: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Health Insurance: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Health Insurance					
	Intent-To-Treat Impact Estimates				
Moderator/Subgroup (Sample N=1,638)	Non-Head Start Group Mean	Impact of Head Start on Subgroup		Difference in Impact Between Subgroups	
		Demographic Covariate Only	With Fall Measure	Demographic Covariate Only	With Fall Measure
Special Needs				0.00	0.02
Special Needs	0.90	0.01	0.04		
No Special Needs	0.88	0.01	0.02		
Parent Married				0.03	0.02
Married	0.82	0.02	0.03		
Not Married	0.93	0.01	0.01		
Child's Race					
White	0.94	-0.01	0.01	0.03 (vs. Black)	0.03 (vs. Black)
Black	0.98	-0.04	-0.03	0.07* (vs. Hispanic)	0.08* (vs. Hispanic)
Hispanic	0.78	0.03	0.05	0.04 (vs. White)	0.04 (vs. White)
Home Language				0.08*	0.07
Not English	0.73	0.05*	0.06*		
English	0.95	-0.02	-0.01		
Depression (Continuous)				0.01	0.01

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.7.2.3: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Dental Care: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Child Has Dental Care					
	Intent-To-Treat Impact Estimates				
		Impact of Head Start on Subgroup		Difference in Impact Between Subgroups	
Moderator/Subgroup (Sample N=1,638)	**Non-Head Start Group Mean**	**Demographic Covariate Only**	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				0.08	0.07
Special Needs	0.47	0.24*	0.19		
No Special Needs	0.52	0.16***	0.12***		
Parent Married				0.02	0.03
Married	0.54	0.18***	0.15**		
Not Married	0.50	0.16***	0.12**		
Child's Race					
White	0.52	0.17***	0.14**	0.07 (vs. Black)	0.08 (vs. Black)
Black	0.52	0.11	0.06	0.12 (vs. Hispanic)	0.13 (vs. Hispanic)
Hispanic	0.51	0.22***	0.19***	0.05 (vs. White)	0.05 (vs. White)
Home Language				0.07	0.08
Not English	0.53	0.22***	0.19***		
English	0.52	0.15***	0.11**		
Depression (Continuous)				0.16***	0.17***

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.7.2.4: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Dental Care: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Child Has Dental Care					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Means	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				0.00	0.04
Special Needs	0.60	0.16	0.17		
No Special Needs	0.57	0.16***	0.13**		
Parent Married				0.04	0.03
Married	0.58	0.18***	0.15**		
Not Married	0.55	0.14**	0.12*		
Child's Race					
White	0.52	0.24***	0.20**	0.11 (vs. Black)	0.11 (vs. Black)
Black	0.59	0.13	0.09	0.01 (vs. Hispanic)	0.01 (vs. Hispanic)
Hispanic	0.60	0.12*	0.10	0.12 (vs. White)	0.10 (vs. White)
Home Language				0.02	0.02
Not English	0.59	0.17**	0.15*		
English	0.56	0.16***	0.13**		
Depression (Continuous)				0.16***	0.16***

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.7.2.5: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Care for Injury in Last Month: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Child Had Care for Injury in Last Month					
	Intent-To-Treat Impact Estimates				
Moderator/Subgroup (Sample N=1,638)	Non-Head Start Group Mean	Impact of Head Start on Subgroup		Difference in Impact Between Subgroups	
		Demographic Covariate Only	With Fall Measure	Demographic Covariate Only	With Fall Measure
Special Needs				0.01	0.01
Special Needs	0.12	0.00	0.01		
No Special Needs	0.08	0.01	0.00		
Parent Married				0.02	0.02
Married	0.09	-0.01	-0.01		
Not Married	0.08	0.01	0.01		
Child's Race					
White	0.04	0.07**	0.07***	0.08* (vs. Black)	0.08* (vs. Black)
Black	0.08	-0.01	-0.01	0.05 (vs. Hispanic)	0.05 (vs. Hispanic)
Hispanic	0.12	-0.06*	-0.06*	0.13** (vs. White)	0.13*** (vs. White)
Home Language				0.02	0.02
Not English	0.09	-0.01	-0.01		
English	0.08	0.01	0.01		
Depression (Continuous)				0.00	0.00

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.7.2.6: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Care for Injury in Last Month: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Child Had Care for Injury in Last Month					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				0.02	0.02
Special Needs	0.18	-0.03	-0.03		
No Special Needs	0.11	-0.00	-0.00		
Parent Married				0.04	0.04
Married	0.08	0.01	0.01		
Not Married	0.15	-0.02	-0.02		
Child's Race					
White	0.10	0.03	0.03	0.06 (vs. Black)	0.06 (vs. Black)
Black	0.19	-0.03	-0.03	0.00 (vs. Hispanic)	0.00 (vs. Hispanic)
Hispanic	0.09	-0.03	-0.03	0.06 (vs. White)	0.06 (vs. White)
Home Language				0.04	0.04
Not English	0.10	-0.04	-0.04		
English	0.13	0.00	0.00		
Depression (Continuous)				0.01	0.01

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.7.2.7: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Need for Ongoing Care: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Child Needs Ongoing Care					
	Intent-To-Treat Impact Estimates				
Moderator/Subgroup (Sample N=1,638)	Non-Head Start Group Mean	Impact of Head Start on Subgroup		Difference in Impact Between Subgroups	
		Demographic Covariate Only	With Fall Measure	Demographic Covariate Only	With Fall Measure
Special Needs				0.11	0.10
Special Needs	0.25	0.09	0.10		
No Special Needs	0.11	-0.01	0.01		
Parent Married				0.02	0.03
Married	0.12	-0.01	0.00		
Not Married	0.13	0.00	0.02		
Child's Race					
White	0.15	-0.02	0.00	0.02 (vs. Black)	0.02 (vs. Black)
Black	0.17	0.00	0.03	0.01 (vs. Hispanic)	0.00 (vs. Hispanic)
Hispanic	0.07	0.02	0.03	0.04 (vs. White)	0.02 (vs. White)
Home Language				0.01	0.02
Not English	0.07	-0.01	0.00		
English	0.15	-0.00	0.02		
Depression (Continuous)				-0.00	0.00

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.7.2.8: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Need for Ongoing Care: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Child Needs Ongoing Care					
	Intent-To-Treat Impact Estimates				
Moderator/Subgroup (Sample N=1,638)	Non-Head Start Group Mean	Impact of Head Start on Subgroup		Difference in Impact Between Subgroups	
		Demographic Covariate Only	With Fall Measure	Demographic Covariate Only	With Fall Measure
Special Needs				0.05	0.02
Special Needs	0.26	-0.04	0.00		
No Special Needs	0.09	0.01	0.03		
Parent Married				0.01	0.01
Married	0.09	0.01	0.02		
Not Married	0.13	-0.00	0.03		
Child's Race					
White	0.13	0.03	0.05	0.03 (vs. Black)	0.02 (vs. Black)
Black	0.14	-0.00	0.03	0.03 (vs. Hispanic)	0.03 (vs. Hispanic)
Hispanic	0.08	-0.03	-0.01	0.06 (vs. White)	0.06 (vs. White)
Home Language				0.04	0.05
Not English	0.06	-0.03	-0.01		
English	0.14	0.01	0.04*		
Depression (Continuous)				0.00	-0.00

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.7.2.9: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Health Status: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Child Health Status					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				0.17	0.15
Special Needs	0.54	0.20*	0.19*		
No Special Needs	0.78	0.04	0.04		
Parent Married				0.06	0.05
Married	0.74	0.09	0.08*		
Not Married	0.77	0.02	0.03		
Child's Race					
White	0.79	0.03	0.02	0.04 (vs. Black)	0.01 (vs. Black)
Black	0.83	-0.01	0.01	0.15* (vs. Hispanic)	0.12 (vs. Hispanic)
Hispanic	0.65	0.14***	0.12**	0.11 (vs. White)	0.10 (vs. White)
Home Language				0.13**	0.12*
Not English	0.62	0.15***	0.14**		
English	0.81	0.02	0.02		
Depression (Continuous)				0.06*	0.05*

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.7.2.10: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Health Status: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Child Health Status					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Special Needs				0.18	0.22*
Special Needs	0.77	-0.20	-0.23*		
No Special Needs	0.82	0.02	-0.01		
Parent Married				0.12*	0.10
Married	0.88	-0.09*	-0.08**		
Not Married	0.75	0.03	0.02		
Child's Race					
White	0.90	-0.04	-0.05	0.07 (vs. Black)	0.08 (vs. Black)
Black	0.82	0.02	0.02	0.09* (vs. Hispanic)	0.08 (vs. Hispanic)
Hispanic	0.74	-0.07	-0.06	0.02 (vs. White)	0.01 (vs. White)
Home Language				0.10*	0.07
Not English	0.76	0.10**	-0.08*		
English	0.84	0.00	-0.01		
Depression (Continuous)				-0.03	-0.03

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Appendix 8.1: Parenting Practices Domain, Estimated Impact of Program Participation

Exhibits A.8.1.1 and A.8.1.2 provide impact estimates that are adjusted for the fact that some of the children assigned to Head Start failed to take advantage of this opportunity—i.e., they never participated in Head Start (the “no shows”)─and that some children assigned to the non-Head Start group managed to find their way into the program. With the exception of one instance (the number of times a child is spanked), all measures of the impact of **participation** in Head Start exceed the corresponding measures of the average impact of **access to** Head Start, regardless of which adjustment method is used. For example, the first row of Exhibit A.8.1.1 shows an estimated impact of 0.17 on the extent of parents’ reading to their child for the average child in the 3-year-old group granted access to the program through assignment to the Head Start research sample. A larger average impact of 0.19-0.20 is shown for those children who actually attended Head Start. These same increases are again seen in the relationship between the calculated effect sizes for the estimates of “access” vs. “participation.”

These data provide an indication of the dampening effect of the crossover children in the non-Head Start group, i.e., reducing the size of any observed difference in outcomes between the Head Start and non-Head Start groups. But, again, the combined no-show/crossover adjustments should only be used as a rough indication of the likely consequences of the presence of crossovers on the interpretation of the impact of Head Start on children’s school readiness.

Exhibit A.8.1.1: Initial Estimates of the Impact of Head Start on Parenting Outcomes, Intent to Treat, and Impact on the Treated: Statistically Significant Results Only, 3-Year-Old Group, Combined English-English and Spanish-English Group (Weighted Data)

Outcome Measure	Impact Estimates			Effect Sizes		
	Impact of Access to Head Start	**Impact of Head Start Participation, No-Show Adjustment**	**Impact of Head Start Participation, Combined No-Show and Crossover Adjustment**	**Impact of Access to Head Start**	**Impact of Head Start Participation, No-Show Adjustment**	**Impact of Head Start Participation, Combined No-Show and Crossover Adjustment**
Overall Impact						
Number of Times Child Is Read To	0.17**	0.19**	0.20**	0.18	0.20	0.22
Family Cultural Enrichment Scale	0.15*	0.17*	0.16*	0.11	0.12	0.11
Spanked Child in Last Week	-0.07*	-0.08*	-0.07	-0.14	-0.16	-0.14
Number Time Spanked Child in Last Week	-0.16*	-0.18*	-0.12	-0.10	-0.11	-0.08
Difference in Impact [1]						
Spanked Child in Last Week (Teen Mom Impact Exceeds Not Teen Mom)	0.16**	0.18**	N/A	0.32	0.36	N/A
Spanked Child in Last Week: Depression	-0.07*	-0.18*	N/A	-0.14	-0.16	N/A
Number of Times Spanked Child: Depression	0.01*	0.01*	N/A	0.01	0.01	N/A
Parental Safety Practices Scale: Home Language (English Impact Exceeds Not English)	0.09*	0.10*	N/A	0.27	0.30	N/A
Safety Devices Subscale (English Impact Exceeds Not English)	0.22*	0.25*	N/A	0.29	0.33	N/A
Impact on Subgroup [2]						
Number of Times Child Is Read To: Not Teen Mom	0.16*	0.18*	N/A	0.17	0.19	N/A
Number of Times Child Is Read To: Female	0.23*	0.25*	N/A	0.25	0.27	N/A
Number of Times Child Is Read To: White	0.27*	0.31*	N/A	0.29	0.33	N/A

Exhibit A.8.1.1: (continued)

Outcome Measure	Impact Estimates			Effect Sizes		
	Impact of Access to Head Start	**Impact of Head Start Participation, No-Show Adjustment**	**Impact of Head Start Participation, Combined No-Show and Crossover Adjustment**	**Impact of Access to Head Start**	**Impact of Head Start Participation, No-Show Adjustment**	**Impact of Head Start Participation, Combined No-Show and Crossover Adjustment**
Number of Times Child Is Read To: Parent Married	0.28**	0.31**	N/A	0.30	0.33	N/A
Number of Times Child Is Read To: Home Language is English	0.19**	0.22*	N/A	0.20	0.24	N/A
Family Cultural Enrichment Scale: Not Teen Mom	0.23**	0.26*	N/A	0.16	0.19	N/A
Family Cultural Enrichment Scale: Male	0.28*	0.33*	N/A	0.20	0.24	N/A
Family Cultural Enrichment Scale: Black	0.24*	0.27*	N/A	0.17	0.19	N/A
Number of Time Outs in Last Week: Female	-0.32*	-0.35*	N/A	-0.17	-0.18	N/A
Spanked Child in Last Week: Teen Mom	-0.17***	-0.19**	N/A	-0.34	-0.38	N/A
Spanked Child in Last Week: Male	-0.11*	-0.12*	N/A	-0.22	-0.24	N/A
Spanked Child in Last Week: Home Language English	-0.10*	-0.11*	N/A	-0.20	-0.22	N/A
Spanked Child in Last Week: Parent Married	-0.11*	-0.13*	N/A	-0.22	-0.26	N/A
Number of Times Spanked Child: Teen Mom	-0.36*	-0.41*	N/A	-0.23	-0.26	N/A
Number of Times Spanked Child: Black	-0.35*	-0.39*	N/A	-0.22	-0.25	N/A
Number of Times Spanked Child: Home Language English	-0.25**	-0.28**	N/A	-0.16	-0.18	N/A

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.

[1] A total of 80 differences in impacts between subgroups were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 8.2. Findings for baseline factors other than depression indicate the amount by which Head Start's estimated impact for the first subset of participants listed in the row label exceeds that for the second subset listed. Findings for depression indicate the change in Head Start's estimated impact that accompanies a 1-point increase in mother's baseline depression score.

[2] A total of 110 subgroup impacts were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 8.2.

Exhibit A.8.1.2: Initial Estimates of the Impact of Head Start on Parenting Outcomes, Intent to Treat, and Impact on the Treated: Statistically Significant Results Only, 4-Year-Old Group, Combined English-English Spanish-English Group (Weighted Data)

Outcome Measure	Impact Estimates			Effect Sizes		
	Impact of Access to Head Start	Impact of Head Start Participation, No-Show Adjustment	Impact of Head Start Participation, Combined No-Show and Crossover Adjustment	Impact of Access to Head Start	Impact of Head Start Participation, No-Show Adjustment	Impact of Head Start Participation, Combined No-Show and Crossover Adjustment
Overall Impact						
Number of Times Child Is Read To	0.13**	0.16**	0.19*	0.13	0.16	0.19
Difference in Impact [1]						
Spanked Child in Last Week: Gender (Female Impact Exceeds Male)	0.15*	0.18*	N/A	0.31	0.37	N/A
Used Time Out in Last Week: Depression	-0.09*	-0.11	N/A	-0.19	-0.23	N/A
Impact on Subgroup [2]						
Number of Times Child Is Read To: Not Teen Mom	0.18*	0.22*	N/A	0.18	0.22	N/A
Family Cultural Enrichment Scale: Hispanic	0.22*	0.25*	N/A	0.15	0.17	N/A
Used Time Out in Last Week: Not Teen Mom	-0.12*	-0.14*	N/A	-0.26	-0.30	N/A
Used Time Out in Last Week: Male	-0.12*	-0.15*	N/A	-0.26	-0.32	N/A
Used Time Out in Last Week: White	-0.11**	-0.13**	N/A	-0.23	-0.28	N/A
Used Time Out in Last Week: Parent Not Married	-0.08*	-0.09*	N/A	-0.17	-0.19	N/A
Used Time Out in Last Week: Home Language English	-0.11**	-0.14**	N/A	-0.23	-0.30	N/A
Safety Devices Subscale: Home Language Not English	0.22*	0.24*	N/A	0.29	0.31	N/A

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.

[1] A total of 80 differences in impacts between subgroups were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 8.2. Findings for depression indicate the change in Head Start's estimated impact that accompanies a 1-point increase in mother's baseline depression score. Findings for baseline factors other than depression indicate the amount by which Head Start's estimated impact for the first subset of participants listed in the row label exceeds that for the second subset listed.

[2] A total of 110 subgroup impacts were examined. The complete set of results, including differences not found to be statistically significant, appears in Appendix 8.2.

Appendix 8.2: Factors That Moderate the Impact of Head Start: Detailed Tables for Parenting Outcomes

The following tables (Exhibits A.8.2.1 through A.8.2.20) provide the results of the moderator/subgroup analyses for measures of parenting outcomes, with a separate table for each individual measure. For clarity, these results are only presented for the full combined sample (i.e., not separately for the English-English and Spanish-English language groups). Each table is organized as follows:

- The first column lists the variable used in the particular subgroup/moderator analysis (separate regressions were estimated for each moderator). For example, analyses were conducted to examine the extent to which child gender was related to program impact.
- As shown, separate lines are shown for the overall construct (e.g., gender) and for each of the subgroups that make up the construct, e.g., boys and girls. Estimates associated with the overall construct represent estimated **differences** in impacts (e.g., boys vs. girls), while the figures associated with each of the subgroup rows represent the **impact on the individual subgroups** (e.g., impacts on boys alone).
- For comparison purposes, the next column provides the mean on the particular outcome measure for the group indicated among children in the non-Head Start group in spring 2003 (the end of the first program year).
- The next set of columns provides the estimated impact on the individual subgroups, while the last two columns provide the estimated difference in impact between subgroups.
- As with the overall impact tables provided in Chapter 8, the estimated impacts are shown using two separate estimation specifications: (1) using regression analyses that include only demographic covariates measured in fall 2002, and (2) using regression analyses that added a measure of the outcome variable assessed in fall 2002. The highlighting indicates which estimate is considered the "best" (see Chapter 4) and which is highlighted in the discussion in Chapter 8.

Exhibit A.8.2.1: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on the Safety Devices Subscale: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Safety Devices Subscale					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Teen Birth				0.12	0.07
Was a Teen Mom	3.15	0.15	0.09		
Was Not a Teen Mom	3.39	0.03	0.02		
Child's Gender				0.03	0.05
Girl	3.3	0.09	0.07		
Boy	3.31	0.05	0.02		
Child's Race					
White	3.32	0.11	0.10	0.01 (vs. Black)	0.06 (vs. Black)
Black	3.19	0.10	0.04	0.10 (vs. Hispanic)	0.03 (vs. Hispanic)
Hispanic	3.39	0.00	0.01	0.11 (vs. White)	0.09 (vs. White)
Depression (Continuous)				0.00	-0.00
Parent Married				0.07	0.05
Married	3.4	0.03	0.02		
Not married	3.22	0.10	0.07		
Home Language				0.26**	0.22*
Not English	3.50	-0.11	-0.11		
English	3.23	0.14*	0.11		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.8.2.2: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on the Safety Devices Subscale: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Safety Devices Subscale					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Teen Birth				0.04	0.03
Was a Teen Mom	3.24	0.09	0.12		
Was Not a Teen Mom	3.42	0.05	0.09		
Child's Gender				0.08	0.05
Girl	3.31	0.10	0.13		
Boy	3.4	0.02	0.08		
Child's Race					
White	3.49	0.05	0.03	0.03 (vs. Black)	0.06 (vs. Black)
Black	3.25	0.08	0.09	0.03 (vs. Hispanic)	0.08 (vs. Hispanic)
Hispanic	3.32	0.06	0.17	0.01 (vs. White)	0.14 (vs. White)
Depression (Continuous)				-0.00	-0.00
Parent Married				0.02	0.06
Married	3.41	0.07	0.06		
Not Married	3.31	0.05	0.12		
Home Language				0.05	0.17
Not English	3.35	0.10	0.22*		
English	3.36	0.04	0.04		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.8.2.3: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Used Time Out in Last Week: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Time Out in Last Week					
	Intent-To-Treat Impact Estimates				
Moderator/Subgroup (Sample N=1,638)	**Non-Head Start Group Mean**	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Teen Birth				0.08	0.07
Was a Teen Mom	0.73	-0.08	-0.07		
Was Not a Teen Mom	0.62	-0.00	0.01		
Child's Gender				0.08	0.08
Girl	0.66	0.01	0.02		
Boy	0.66	-0.07	-0.06		
Child's Race					
White	0.76	-0.07	-0.06	0.11 (vs. Black)	0.10 (vs. Black)
Black	0.61	0.03	0.04	0.09 (vs. Hispanic)	0.08 (vs. Hispanic)
Hispanic	0.61	-0.06	-0.04	0.02 (vs. White)	0.02 (vs. White)
Depression (Continuous)				-0.03	-0.03
Parent Married				0.03	0.01
Married	0.67	-0.04	-0.02		
Not Married	0.65	-0.02	-0.02		
Home Language				0.10	0.11
Not English	0.54	-0.11	-0.10		
English	0.70	-0.01	0.01		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.8.2.4: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Used Time Out in Last Week, 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Time Out in Last Week					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Teen Birth				0.05	0.06
Was a Teen Mom	0.68	-0.06	-0.06		
Was Not a Teen Mom	0.68	-0.10	-0.12*		
Child's Gender				0.03	0.05
Girl	0.72	-0.07	-0.07		
Boy	0.64	-0.10*	-0.12*		
Child's Race					
White	0.81	-0.10*	-0.11**	0.03 (vs. Black)	0.00 (vs. Black)
Black	0.67	-0.12	-0.11	0.06 (vs. Hispanic)	0.04 (vs. Hispanic)
Hispanic	0.59	-0.06	-0.08	0.04 (vs. White)	0.04 (vs. White)
Depression (Continuous)				-0.09*	-0.09*
Parent Married				0.00	0.03
Married	0.65	-0.09	-0.11		
Not Married	0.71	-0.08	-0.08*		
Home Language				0.07	0.07
Not English	0.52	-0.03	-0.04		
English	0.77	-0.10**	-0.11**		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.8.2.5: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Child Spanked in Last Week 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Child Spanked in Last Week					
Moderator/Subgroup (Sample N=1,638)	Intent-To-Treat Impact Estimates				
	Non-Head Start Group Mean	Impact of Head Start on Subgroup		Difference in Impact Between Subgroups	
		Demographic Covariate Only	With Fall Measure	Demographic Covariate Only	With Fall Measure
Teen Birth				0.16**	0.17**
Was a Teen Mom	0.56	-0.17***	-0.17**		
Was Not a Teen Mom	0.44	-0.01	0.00		
Child's Gender				0.08	0.09
Girl	0.48	-0.03	-0.01		
Boy	0.49	-0.11*	-0.11		
Child's Race					
White	0.46	-0.04	-0.05	0.05 (vs. Black)	0.02 (vs. Black)
Black	0.54	-0.09	-0.07	0.02 (vs. Hispanic)	0.01 (vs. Hispanic)
Hispanic	0.46	-0.08	-0.06	0.04 (vs. White)	0.01 (vs. White)
Depression (Continuous)				-0.07*	-0.07*
Parent Married				0.07	0.07
Married	0.48	-0.11*	-0.10*		
Not Married	0.48	-0.04	-0.03		
Home Language				0.08	0.07
Not English	0.40	-0.02	-0.01		
English	0.52	-0.10**	-0.08*		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.8.2.6: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Child Spanked in Last Week 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Child Spanked in Last Week					
Moderator/Subgroup (Sample N=1,638)	Intent-To-Treat Impact Estimates				
	Non-Head Start Group Mean	Impact of Head Start on Subgroup		Difference in Impact Between Subgroups	
		Demographic Covariate Only	With Fall Measure	Demographic Covariate Only	With Fall Measure
Teen Birth				0.08	0.08
Was a Teen Mom	0.34	0.05	0.05		
Was Not a Teen Mom	0.39	-0.04	-0.04		
Child's Gender				0.14	0.15*
Girl	0.41	-0.07	-0.08		
Boy	0.34	0.07	0.07		
Child's Race					
White	0.34	-0.04	-0.02	0.04 (vs. Black)	0.00 (vs. Black)
Black	0.43	0.00	-0.01	0.01 (vs. Hispanic)	0.02 (vs. Hispanic)
Hispanic	0.37	0.02	0.00	0.05 (vs. White)	0.02 (vs. White)
Depression (Continuous)				-0.01	-0.01
Parent Married				-0.06	-0.06
Married	0.35	0.03	0.02		
Not Married	0.42	-0.03	-0.03		
Home Language				0.00	0.02
Not English	0.37	-0.00	-0.02		
English	0.37	-0.01	-0.00		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.8.2.7: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Restricting Child Movement Scale: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Restricting Child Movement Scale					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Teen Birth				0.05	0.07
Was a Teen Mom	3.84	0.02	0.03		
Was Not a Teen Mom	3.93	-0.03	-0.04		
Child's Gender				0.04	0.03
Girl	3.88	0.00	-0.00		
Boy	3.90	-0.03	-0.03		
Child's Race					
White	3.90	-0.03	-0.02	0.00 (vs. Black)	0.02 (vs. Black)
Black	3.84	-0.03	-0.04	0.04 (vs. Hispanic)	0.06 (vs. Hispanic)
Hispanic	3.94	0.01	0.02	0.04 (vs. White)	0.04 (vs. White)
Depression (Continuous)				0.00	0.00
Parent Married				0.02	0.02
Married	3.92	-0.01	-0.01		
Not Married	3.87	-0.03	-0.03		
Home Language				0.04	0.05
Not English	3.93	0.01	0.02		
English	3.88	-0.03	-0.03		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.8.2.8: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Restricting Child Movement Scale: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Restricting Child Movement Scale					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Teen Birth				0.06	0.07
Was a Teen Mom	3.86	0.02	-0.02		
Was Not a Teen Mom	3.88	0.05	0.04		
Child's Gender				0.00	0.02
Girl	3.87	0.02	0.01		
Boy	3.86	0.03	0.03		
Child's Race					
White	3.85	0.02	0.01	0.04 (vs. Black)	0.04 (vs. Black)
Black	3.82	0.05	0.05	0.04 (vs. Hispanic)	0.03 (vs. Hispanic)
Hispanic	3.91	0.01	0.01	0.01 (vs. White)	0.01 (vs. White)
Depression (Continuous)				-0.00	0.00
Parent Married				0.06	0.04
Married	3.90	-0.00	-0.00		
Not Married	3.85	0.05	0.04		
Home Language				0.03	0.02
Not English	3.93	0.00	0.01		
English	3.84	0.03	0.03		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.8.2.9: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Number of Times Child Had Time Out in Last Week 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Number of Times Had Time Out in Last Week					
Moderator/Subgroup (Sample N=1,638)	Intent-To-Treat Impact Estimates				
	Non-Head Start Group Mean	Impact of Head Start on Subgroup		Difference in Impact Between Subgroups	
		Demographic Covariate Only	With Fall Measure	Demographic Covariate Only	With Fall Measure
Teen Birth				0.01	0.13
Was a Teen Mom	2.14	-0.24	-0.13		
Was Not a Teen Mom	1.76	-0.23	-0.25		
Child's Gender				0.24	0.23
Girl	1.84	-0.34*	-0.32*		
Boy	1.98	-0.11	-0.09		
Child's Race					
White	2.45	-0.32	-0.29	0.04 (vs. Black)	0.02 (vs. Black)
Black	1.98	-0.28	-0.26	0.19 (vs. Hispanic)	0.19 (vs. Hispanic)
Hispanic	1.27	-0.09	-0.08	0.24 (vs. White)	0.21 (vs. White)
Depression (Continuous)				0.00	0.00
Parent Married				0.17	0.11
Married	1.90	-0.28	-0.24		
Not Married	1.91	-0.11	-0.12		
Home Language				0.23	0.21
Not English	0.88	-0.09	-0.08		
English	2.30	-0.32	-0.28		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.8.2.10: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Number of Times Child Had Time Out in Last Week: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Number of Times Had Time Out in Last Week					
Moderator/Subgroup (Sample N=1,638)	Intent-To-Treat Impact Estimates				
	Non-Head Start Group Mean	Impact of Head Start on Subgroup		Difference in Impact Between Subgroups	
		Demographic Covariate Only	With Fall Measure	Demographic Covariate Only	With Fall Measure
Teen Birth				0.46	0.53
Was a Teen Mom	1.73	0.33	0.38		
Was Not a Teen Mom	1.62	-0.13	-0.15		
Child's Gender				0.31	0.16
Girl	1.84	-0.13	-0.04		
Boy	1.45	0.19	0.12		
Child's Race					
White	2.14	0.44	0.36	0.91* (vs. Black)	0.59 (vs. Black)
Black	1.93	-0.47	-0.22	0.48 (vs. Hispanic)	0.17 (vs. Hispanic)
Hispanic	1.13	0.01	-0.05	0.44 (vs. White)	0.42 (vs. White)
Depression (Continuous)				0.00	0.00
Parent Married				0.17	0.13
Married	1.50	-0.07	-0.02		
Not Married	1.79	0.09	0.11		
Home Language				0.15	0.21
Not English	0.86	-0.04	-0.09		
English	2.06	0.10	0.13		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.8.2.11: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Number of Times Read to: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Number of Times Read to					
	Intent-To-Treat Impact Estimates				
Moderator/Subgroup (Sample N=1,638)	Non-Head Start Group Mean	Impact of Head Start on Subgroup		Difference in Impact Between Subgroups	
		Demographic Covariate Only	With Fall Measure	Demographic Covariate Only	With Fall Measure
Teen Birth				0.02	0.04
Was a Teen Mom	2.68	0.18	0.16		
Was Not a Teen Mom	2.83	0.16*	0.12*		
Child's Gender				0.13	0.02
Girl	2.81	0.23*	0.14		
Boy	2.73	0.10	0.12		
Child's Race					
White	2.90	0.27*	0.26*	0.10 (vs. Black)	0.16 (vs. Black)
Black	2.71	0.16	0.10	0.09 (vs. Hispanic)	0.06 (vs. Hispanic)
Hispanic	2.7	0.07	0.07	0.20 (vs. White)	0.22 (vs. White)
Depression (Continuous)				-0.00	-0.00
Parent Married				0.19	0.17
Married	2.77	0.28**	0.24*		
Not Married	2.78	0.10	0.07		
Home Language				0.13	0.16
Not English	2.69	0.06	0.01		
English	2.81	0.19**	0.17**		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.8.2.12: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Number of Times Read to: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Number of Times Read to					
Moderator/Subgroup (Sample N=1,638)	Intent-To-Treat Impact Estimates				
	Non-Head Start Group Mean	Impact of Head Start on Subgroup		Difference in Impact Between Subgroups	
		Demographic Covariate Only	With Fall Measure	Demographic Covariate Only	With Fall Measure
Teen Birth				0.13	0.13
Was a Teen Mom	2.85	0.05	0.04		
Was Not a Teen Mom	2.78	0.18*	0.16*		
Child's Gender				0.13	0.13
Girl	2.74	0.07	0.05		
Boy	2.87	0.20	0.18		
Child's Race					
White	3.13	0.12	0.09	0.14 (vs. Black)	0.13 (vs. Black)
Black	2.89	-0.01	-0.05	0.25 (vs. Hispanic)	0.29 (vs. Hispanic)
Hispanic	2.50	0.23	0.24	0.11 (vs. White)	0.15 (vs. White)
Depression (Continuous)				-0.00	-0.00
Parent Married				0.02	0.03
Married	2.84	0.11	0.09		
Not Married	2.76	0.13	0.12		
Home Language				0.09	0.13
Not English	2.45	0.20	0.21		
English	2.98	0.11	0.08		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.8.2.13: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Parental Safety Practices Scale: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Parental Safety Practices Scale					
Moderator/Subgroup (Sample N=1,638)	Intent-To-Treat Impact Estimates				
	Non-Head Start Group Mean	Impact of Head Start on Subgroup		Difference in Impact Between Subgroups	
		Demographic Covariate Only	With Fall Measure	Demographic Covariate Only	With Fall Measure
Teen Birth				0.09	0.07
Was a Teen Mom	3.60	0.08	0.06		
Was Not a Teen Mom	3.75	-0.00	-0.01		
Child's Gender				0.03	0.03
Girl	3.70	0.04	0.03		
Boy	3.69	0.02	0.00		
Child's Race					
White	3.69	0.04	0.04	0.01 (vs. Black)	0.04 (vs. Black)
Black	3.65	0.03	0.00	0.01 (vs. Hispanic)	0.01 (vs. Hispanic)
Hispanic	3.74	0.02	0.01	0.02 (vs. White)	0.03 (vs. White)
Depression (Continuous)				0.00	-0.00
Parent Married				0.03	0.02
Married	3.74	0.01	0.00		
Not Married	3.65	0.04	0.02		
Home Language				0.09*	0.07
Not English	3.78	-0.04	-0.04		
English	3.66	0.05	0.04		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.8.2.14: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Parental Safety Practices Scale: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Parental Safety Practices Scale					
Moderator/Subgroup (Sample N=1,638)	Intent-To-Treat Impact Estimates				
	Non-Head Start Group Mean	Impact of Head Start on Subgroup		Difference in Impact Between Subgroups	
		Demographic Covariate Only	With Fall Measure	Demographic Covariate Only	With Fall Measure
Teen Birth				0.00	0.01
Was a Teen Mom	3.66	0.03	0.03		
Was Not a Teen Mom	3.73	0.03	0.04		
Child's Gender				0.03	0.02
Girl	3.69	0.05	0.05		
Boy	3.72	0.01	0.03		
Child's Race					
White	3.72	0.02	0.00	0.03 (vs. Black)	0.04 (vs. Black)
Black	3.67	0.05	0.04	0.02 (vs. Hispanic)	0.03 (vs. Hispanic)
Hispanic	3.7	0.03	0.07	0.01 (vs. White)	0.07 (vs. White)
Depression (Continuous)				-0.00	-0.00
Parent Married				0.04	0.05
Married	3.74	0.01	0.01		
Not Married	3.68	0.05	0.06		
Home Language				0.01	0.06
Not English	3.73	0.02	0.08		
English	3.69	0.03	0.02		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.8.2.15: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Number of Times Spanked in Last Week : 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Number of Times Spanked in Last Week					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Teen Birth				0.32	0.23
Was a Teen Mom	1.19	-0.36*	-0.21		
Was Not a Teen Mom	0.87	-0.05	0.03		
Child's Gender				0.07	0.05
Girl	0.92	-0.20	-0.03		
Boy	1.06	-0.13	-0.08		
Child's Race					
White	0.93	-0.06	0.02	0.29 (vs. Black)	0.24 (vs. Black)
Black	1.27	-0.35*	-0.21	0.27 (vs. Hispanic)	0.23 (vs. Hispanic)
Hispanic	0.76	-0.08	0.02	0.02 (vs. White)	0.00 (vs. White)
Depression (Continuous)				0.01*	0.01*
Parent Married				0.05	0.03
Married	0.91	-0.16	-0.06		
Not Married	1.05	-0.10	-0.03		
Home Language				0.27	0.11
Not English	0.59	0.03	0.02		
English	1.14	-0.25**	-0.09		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.8.2.16: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Number of Times Spanked in Last Week : 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Number of Times Spanked in Last Week					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Teen Birth				0.21	0.25
Was a Teen Mom	0.71	0.15	0.20		
Was Not a Teen Mom	0.67	-0.06	-0.05		
Child's Gender				0.08	0.06
Girl	0.60	0.06	0.07		
Boy	0.76	-0.02	0.01		
Child's Race					
White	0.64	0.06	0.13	0.11 (vs. Black)	0.18 (vs. Black)
Black	0.88	-0.05	-0.05	0.06 (vs. Hispanic)	0.06 (vs. Hispanic)
Hispanic	0.61	0.02	0.01	0.04 (vs. White)	0.12 (vs. White)
Depression (Continuous)				-0.00	-0.00
Parent Married				0.02	0.02
Married	0.58	0.01	0.01		
Not Married	0.77	-0.02	0.03		
Home Language				0.01	0.07
Not English	0.58	0.02	-0.01		
English	0.73	0.02	0.06		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.8.2.17: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Family Cultural Enrichment Scale: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Family Cultural Enrichment Scale					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Teen Birth				0.12	0.21
Was a Teen Mom	3.43	0.12	0.02		
Was Not a Teen Mom	3.60	0.23*	0.23**		
Child's Gender				0.15	0.25
Girl	3.56	0.12	0.03		
Boy	3.52	0.27*	0.28*		
Child's Race					
White	3.25	0.13	0.10	0.17 (vs. Black)	0.13 (vs. Black)
Black	3.63	0.30*	0.24*	0.17 (vs. Hispanic)	0.12 (vs. Hispanic)
Hispanic	3.73	0.13	0.11	0.00 (vs. White)	0.01 (vs. White)
Depression (Continued)				0.00	0.00
Parent Married				0.15	0.10
Married	3.47	0.25*	0.18		
Not Married	3.6	0.11	0.09		
Home Language				0.15	0.28
Not English	3.45	0.30	0.35		
English	3.57	0.15	0.07		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.8.2.18: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Family Cultural Enrichment Scale: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Family Cultural Enrichment Scale					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Means	**Impact of Head Start on Subgroup**		**Difference in impact between adjacent subgroups**	
		Demographic covariate only	**With fall measure**	**Demographic covariate only**	**With fall measure**
Teen Birth				0.07	0.04
Was a Teen Mom	3.81	0.03	0.08		
Was Not a Teen Mom	3.96	0.11	0.12		
Child's Gender				0.31	0.23
Girl	3.79	0.24	0.22		
Boy	4.01	-0.07	-0.01		
Child's Race					
White	3.93	-0.01	0.07	0.09 (vs. Black)	0.11 (vs. Black)
Black	4.30	0.07	0.04	0.09 (vs. Hispanic)	0.26 (vs. Hispanic)
Hispanic	3.67	0.16	0.22*	0.18 (vs. White)	0.15 (vs. White)
Depression (Continuous)				0.00	0.00
Parent Married				0.01	0.05
Married	3.90	0.06	0.05		
Not Married	3.91	0.07	0.10		
Home Language				0.01	0.04
Not English	3.65	0.08	0.14		
English	4.04	0.09	0.10		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.8.2.19: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Removing Harmful Objects Scale: 3-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Removing Harmful Objects Scale					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Teen Birth				0.09	0.07
Was a Teen Mom	3.80	0.08	0.07		
Was Not a Teen Mom	3.92	-0.01	-0.00		
Child's Gender				0.01	0.01
Girl	3.87	0.03	0.03		
Boy	3.88	0.02	0.02		
Child's Race					
White	3.83	0.03	0.04	0.01 (vs. Black)	0.03 (vs. Black)
Black	3.91	0.02	0.02	0.00 (vs. Hispanic)	0.01 (vs. Hispanic)
Hispanic	3.89	0.02	0.01	0.01 (vs. White)	0.03 (vs. White)
Depression (Continuous)				0.00	0.00
Parent Married				0.03	0.04
Married	3.89	0.01	0.00		
Not Married	3.86	0.05	0.04		
Home Language				0.06	0.07
Not English	3.93	-0.02	-0.03		
English	3.85	0.04	0.04		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.

Exhibit A.8.2.20: Initial One-Year Estimates of Factors That Moderate Head Start's Impact on Removing Harmful Objects Scale: 4-Year-Old Group, Combined Fall English-Spring English and Fall Spanish-Spring English Group, Weighted Data

Removing Harmful Objects Scale					
Moderator/Subgroup (Sample N=1,638)	**Intent-To-Treat Impact Estimates**				
	Non-Head Start Group Mean	**Impact of Head Start on Subgroup**		**Difference in Impact Between Subgroups**	
		Demographic Covariate Only	**With Fall Measure**	**Demographic Covariate Only**	**With Fall Measure**
Teen Birth				0.01	0.00
Was a Teen Mom	3.88	0.01	-0.00		
Was Not a Teen Mom	3.89	-0.00	-0.00		
Child's Gender				0.02	0.02
Girl	3.87	0.01	0.01		
Boy	3.90	-0.01	-0.01		
Child's Race					
White	3.86	-0.00	-0.02	-0.00 (vs. Black)	0.01 (vs. Black)
Black	3.93	-0.00	-0.01	0.01 (vs. Hispanic)	0.02 (vs. Hispanic)
Hispanic	3.88	0.00	0.02	-0.01 (vs. White)	0.03 (vs. White)
Depression (Continuous)				-0.00	-0.00
Parent Married				0.09	0.07
Married	3.90	-0.04	-0.04		
Not Married	3.87	0.05	0.04		
Home Language				0.03	0.01
Not English	3.91	-0.02	-0.01		
English	3.87	0.01	0.00		

* = p≤0.05, ** = p≤0.01, *** = p≤0.001.