Third Grade Follow-up to the Head Start Impact Study

Final Report

OPRE Report 2012-45

October 2012

THIRD GRADE FOLLOW-UP TO THE HEAD START IMPACT STUDY

FINAL REPORT

OPRE Report 2012-45

October 2012

Michael Puma, Chesapeake Research Associates, Stephen Bell, Abt Associates, Ronna Cook, Ronna Cook Associates, Camilla Heid, Pam Broene, and Frank Jenkins, Westat, Andrew Mashburn, Portland State University, and Jason Downer, University of Virginia

Submitted to:
Jennifer Brooks, Project Officer
Office of Planning, Research and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

Contract Number: HHSP23320062929YC

Project Director: Camilla Heid
Westat
1600 Research Boulevard
Rockville, MD 20850

This report is in the public domain. Permission to reproduce is not necessary. Suggested citation: Mike Puma, Stephen Bell, Ronna Cook, Camilla Heid, Pam Broene, Frank Jenkins, Andrew Mashburn, and Jason Downer (2012). *Third Grade Follow-up to the Head Start Impact Study Final Report,* OPRE Report # 2012-45, Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Disclaimer
The views expressed in this publication do not necessarily reflect the views or policies of the Office of Planning, Research and Evaluation, the Administration for Children and Families, or the U.S. Department of Health and Human Services.

This report and other reports sponsored by the Office of Planning, Research and Evaluation are available at http://www.acf.hhs.gov/programs/opre.

Acknowledgements

This report of the Third Grade Follow-Up to the Head Start Impact Study is the result of several years of design, data collection, and analysis. We gratefully acknowledge the contributions and dedication of individuals and organizations in the preparation and production of this report. A special thanks to Dr. Jennifer Brooks, the Federal Project Officer, for her expert leadership and vision.

There were those who were worried that random assignment and subsequent data collection efforts would be difficult, if not impossible to implement. Study staff has 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, as well as the study children's elementary schools and their staff were instrumental in the successful implementation of this study. The ongoing backing of the Office of Head Start and Regional Office staff was 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 during the data collection 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 the design and analysis presented in the report.

Finally, we gratefully acknowledge the staff from Westat, Ronna Cook Associates, Chesapeake Research Associates, Abt Associates, and the University of Virginia for their hard work, professionalism, and dedication to the project. Particularly, we thank the field staff, data processing staff, and programmers who worked extensively to successfully complete the data collection. A special thanks to Ban Cheah who worked tirelessly to complete the many analysis requests.

Table of Contents

Chapter		Page
	Acknowledgements	i
	Executive Summary	xiii
1	Study Goals and Purposes	1
	The Head Start Program	1
	The Congressional Mandate for This Study	4
	Study Objectives and Research Questions	5
	Contents of Report	8
2	Study Design and Implementation	11
	Study Design Overview	11
	Sample Selection	12
	Random Assignment	17
	The Success of Random Assignment	19
	Data Collection and Data Sources	21
	Defining and Balancing the Analysis Sample	44
	Data Analysis	46
3	School Experiences	55
	Introduction	55
	Measuring 3rd Grade School Experiences	55
	Results	57
	Summary	59
4	The Impact of Head Start on Child and Parent Outcomes at the End of 3rd Grade	75
	Introduction	75
	Intent to Treat (ITT) Impacts	76
	Estimated Impacts on Participants (IOT Impacts)	88
	3rd Grade Findings Within the Context of Earlier Head Start Impact Study Findings	91
	Summary	103

Contents (continued)

Chapter		Page
5	Variation in Impacts: Child and Parent Subgroups	105
	Introduction	105
	Subgroups Examined	106
	Analytic Approach	108
	End of 3rd Grade Results	109
	Impacts That Are Sustained Through 3rd Grade	132
	Summary of Findings	143
6	Conclusions	145
	Introduction	145
	How Does Gaining Access to Head Start Affect Children's Head Start and Early Elementary School Experiences?	146
	What Is the Impact of Head Start on Children and Families?	147
	Is There Variation in Impacts on Children and Families?	149
	Discussion	150
References		155

Appendixes

A	Updated Weighting and Analysis Procedures	A-1
B	Baseline Characteristics for the 3rd Grade Analysis Sample	B-1
C	Intent to Treat (ITT) Tables, 2002-2008	C-1
D	Impact on the Treated (IOT) Tables, 2003-2008	D-1
E	Subgroup Tables, 2003-2008	E-1

Exhibits

1	Percentage of Children by Racial/Ethnic Characteristics and By Age Cohort	xiv
2a	Summary of ITT Cognitive Impacts for 4-Year-Olds by Year	xxiii

Contents (continued)

Exhibits (continued)		Page
2b	Summary of ITT Cognitive Impacts for 3-Year-Olds by Year	xxiv
3a	Summary of ITT Social-Emotional Impacts for 4-Year-Olds by Year	xxvii
3b	Summary of ITT Social-Emotional Impacts for 3-Year-Olds by Year	xxviii
4a	Summary of ITT Health Impacts for 4-Year-Olds by Year	xxx
4b	Summary of ITT Health Impacts for 3-Year-Olds by Year	xxx
5a	Summary of ITT Parenting Practices Impacts for 4-Year-Olds by Year	xxxii
5b	Summary of ITT Parenting Practices Impacts for 3-Year-Olds by Year	xxxiii
2.1	Sample Selection Process for the Head Start Impact Study	16
2.2	Number of Children Randomly Assigned to Head Start and Control Groups, by Age Cohort	19
2.3	The Incidence of No-Show and Crossover Behavior for the Sample as Randomly Assigned, by Age Cohort (Weighted Data)	21
2.4	Treatment and Control Response Rates for All Study Years and for All Study Instruments for the 4-Year-Old Cohort	23
2.5	Treatment and Control Response Rates for All Study Years and for All Study Instruments for the 3-Year-Old Cohort	24
2.6	Cognitive Domain Measures From Direct Assessment and Teacher Report and Year That Measure Was Administered	27
2.7	Social-Emotional, Health, and Parenting Practices Domain Measures From Parent and Teacher Report and Year That Measure Was Administered	29
2.8	Baseline Demographic Variables Included in the Statistical Models Estimating the Impact of Head Start	47
2.9	Factors Used To Define Subgroups	50

3.1	Estimated Impacts on School Experiences at 3rd Grade: 3-Year-Old Cohort	61
3.2	Estimated Impacts on School Experiences at 3rd Grade: 4-Year-Old Cohort	68
4.1	Estimated Impacts on 3rd Grade Cognitive Outcomes: 4-Year-Old Cohort	77
4.2	Estimated Impacts on 3rd Grade Cognitive Outcomes: 3-Year-Old Cohort	78
4.3	Estimated Impacts on 3rd Grade Social-Emotional Outcomes: 4-Year-Old Cohort	81
4.4	Estimated Impacts on 3rd Grade Social-Emotional Outcomes: 3-Year-Old Cohort	83
4.5	Estimated Impacts on 3rd Grade Child Health Outcomes: 4-Year-Old Cohort	85
4.6	Estimated Impacts on 3rd Grade Child Health Outcomes: 3-Year-Old Cohort	86
4.7	Estimated Impacts on 3rd Grade Parenting Practices Outcomes: 4-Year-Old Cohort	87
4.8	Estimated Impacts on 3rd Grade Parenting Practices Outcomes: 3-Year-Old Cohort	88
4.9	Estimated IOT Impacts for 3rd Grade Outcomes Showing Significant ITT Impacts: 4-Year-Old Cohort	90
4.10	Estimated IOT Impacts for 3rd Grade Outcomes Showing Significant ITT Impacts: 3-Year-Old Cohort	91
4.11	ITT Impacts on the Woodcock-Johnson III Letter-Word Identification Test, by Year and Age Cohort	93
4.12	Summary of ITT Cognitive Impacts for 4-Year-Olds by Year	95
4.13	Summary of ITT Cognitive Impacts for 3-Year-Olds by Year	96
4.14	Summary of ITT Social-Emotional Impacts for 4-Year-Olds by Year	98
4.15	Summary of ITT Social-Emotional Impacts for 3-Year-Olds by Year	99

Contents (continued)

Exhibits (continued)		Page
4.16	Summary of ITT Health Impacts for 4-Year-Olds by Year	100
4.17	Summary of ITT Health Impacts for 3-Year-Olds by Year	100
4.18	Summary of ITT Parenting Practices Impacts for 4-Year-Olds by Year	102
4.19	Summary of ITT Parenting Practices Impacts for 3-Year-Olds by Year	103
5.1	Distribution of Children Across the Subgroups by Age Cohort and Random Assignment Status	107
5.2	Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Child's Pre-Academic Skills	110
5.3	Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Child's Pre-Academic Skills	111
5.4	Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Child's Special Needs	112
5.5	Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Child's Special Needs	113
5.6	Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Child's Home Language	114
5.7	Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Child's Home Language	115
5.8	Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Biological Mother/ Caregiver Race/Ethnicity	117
5.9	Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Biological Mother/ Caregiver Race/Ethnicity	118
5.10	Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Parent/Caregiver Reported Depressive Symptoms	122
5.11	Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Parent/Caregiver Reported Depressive Symptoms	124

Contents (continued)

Exhibits (continued)		Page
5.12	Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Household Risk Index	127
5.13	Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Household Risk Index	128
5.14	Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Urbanicity	131
5.15	Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Urbanicity	132
5.16	Summary of the Subgroup Findings for the 4-Year-Old Cohort	133
5.17	Summary of the Subgroup Findings for the 3-Year-Old Cohort	136
A.1	Head Start Impact Study Data Collection	A-2
A.2	Cross-sectional Weights Produced for 3rd Grade Head Start Follow-up Analysis	A-2
A.3	Variables Used to Form Nonresponse Adjustment Cells within Treatment, Control Groups	A-4
A.4	Response Rates for 3rd Grade Follow-up to the Head Start Impact Study	A-5
A.5	Fall 2002 Pretest Measures Found Appropriate for Inclusion as Covariates in 3rd Grade Impact Regressions, by Age Cohort	A-9
A.6	Pretest Measures Used in 3rd Grade Impact Analyses, by Outcome Measure	A-10
B.1A	Covariate Comparison of 3rd Grade Sample: 4-Year-Old Cohort (Child Assessment)	B-1
B.1B	Covariate Comparison of 3rd Grade Sample: 4-Year-Old Cohort (Math Teacher Survey-Teacher Child Report)	B-3
B.1C	Covariate Comparison of 3rd Grade Sample: 4-Year-Old Cohort (Parent Interview)	B-5

Contents (continued)

Exhibits (continued)		Page
B.1D	Covariate Comparison of 3rd Grade Sample: 4-Year-Old Cohort (Principal Survey)	B-7
B.1E	Covariate Comparison of 3rd Grade Sample: 4-Year-Old Cohort (Reading Teacher Survey-Teacher Child Report)	B-9
B.1F	Covariate Comparison of 3rd Grade Sample: 4-Year-Old Cohort (General Teacher Survey-Teacher Child Report)	B-11
B.2A	Covariate Comparison of 3rd Grade Sample: 3-Year-Old Cohort (Child Assessment)	B-13
B.2B	Covariate Comparison of 3rd Grade Sample: 3-Year-Old Cohort (Math Teacher Survey-Teacher Child Report)	B-15
B.2C	Covariate Comparison of 3rd Grade Sample: 3-Year-Old Cohort (Parent Interview)	B-17
B.2D	Covariate Comparison of 3rd Grade Sample: 3-Year-Old Cohort (Principal Survey)	B-19
B.2E	Covariate Comparison of 3rd Grade Sample: 3-Year-Old Cohort (Reading Teacher Survey-Teacher Child Report)	B-21
B.2F	Covariate Comparison of 3rd Grade Sample: 3-Year-Old Cohort (General Teacher Survey-Teacher Child Report)	B-23
C.1A	Estimated Impacts on Cognitive Outcomes by Year: 4-Year-Old Cohort	C-1
C.1B	Estimated Impacts on Cognitive Outcomes by Year: 3-Year-Old Cohort	C-6
C.2A	Estimated Impacts on Social-Emotional Outcomes by Year: 4-Year-Old Cohort	C-12
C.2B	Estimated Impacts on Social-Emotional Outcomes by Year: 3-Year-Old Cohort	C-16
C.3A	Estimated Impacts on Parent-Reported Child Health Outcomes by Year: 4-Year-Old Cohort	C-21

Contents (continued)

Exhibits (continued)		Page
C.3B	Estimated Impacts on Child Health Outcomes by Year: 3-Year-Old Cohort	C-24
C.4A	Estimated Impacts on Parenting Outcomes by Year: 4-Year-Old Cohort	C-27
C.4B	Estimated Impacts on Parenting Outcomes by Year: 3-Year-Old Cohort	C-30
D.1A	Estimated IOT Impacts on Cognitive Outcomes, by Year: 4-Year-Old Cohort	D-1
D.1B	Estimated IOT Impacts on Cognitive Outcomes, by Year: 3-Year-Old Cohort	D-4
D.2A	Estimated IOT Impacts on Social-Emotional Outcomes, by Year: 4-Year-Old Cohort	D-8
D.2B	Estimated IOT Impacts on Social-Emotional Outcomes, by Year: 3-Year-Old Cohort	D-11
D.3A	Estimated IOT Impacts on Parent-Reported Health Outcomes, by Year: 4-Year-Old Cohort	D-14
D.3B	Estimated IOT Impacts on Parent-Reported Health Outcomes, by Year: 3-Year-Old Cohort	D-16
D.4A	Estimated IOT Impacts on Parenting Outcomes, by Year: 4-Year-Old Cohort	D-18
D.4B	Estimated IOT Impacts on Parenting Outcomes, by Year: 3-Year-Old Cohort	D-20
E.1	Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Child's Pre-Academic Skills	E-1
E.2	Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Child's Home Language	E-3
E.3	Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Special Needs	E-5

Contents (continued)

Exhibits (continued)		Page
E.4	Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Biological Mother/ Caregiver Race/Ethnicity	E-7
E.5	Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Parent/Caregiver Reported Depressive Symptoms	E-10
E.6	Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Household Risk Index	E-14
E.7	Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Urbanicity	E-17
E.8	Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Child's Pre-Academic Skills	E-19
E.9	Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Child's Home Language	E-22
E.10	Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Special Needs	E-25
E.11	Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Biological Mother/ Caregiver Race/Ethnicity	E-28
E.12	Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Parent/Caregiver Reported Depressive Symptoms	E-32
E.13	Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Household Risk Index	E-39
E.14	Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Urbanicity	E-43

Third Grade Follow-up to the Head Start Impact Study

Executive Summary

Introduction

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.

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:[1]

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.

- "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?"

The *Head Start Impact Study Final Report* [2]addressed these questions by reporting on the impacts of Head Start on children and families during the children's preschool, kindergarten, and 1st grade years. This *Third Grade Follow-up to the Head Start Impact Study Final Report* addresses these same questions by looking at longer-term effects through the end of 3rd grade.

Background for the Head Start Impact Study

The Head Start Impact Study (HSIS) was conducted with a nationally representative sample of 84 grantee/delegate agencies and included nearly 5,000 newly entering, eligible 3- and

[1] U.S. Department of Health and Human Services, Administration for Children, Youth, and Families. (January, 1999). *Evaluating Head Start: A recommended framework for studying the impact of the Head Start program.* Washington, DC: Author.

[2] U.S. Department of Health and Human Services, Administration for Children and Families (January, 2010). *Head Start Impact Study: Final Report.* Washington, DC. http://www.acf.hhs.gov/programs/opre/hs/impact_study/reports/impact_study/hs_impact_study_final.pdf

4-year-old children who were randomly assigned to either: (1) a Head Start group that had access to Head Start program services or (2) a control group that did not have access to Head Start, but could enroll in other early childhood programs or non-Head Start services selected by their parents. Data collection began in fall 2002 and continued through 2008, following children from program application through the spring of their 3rd grade year.

The study was designed to separately examine two cohorts of children, newly entering 3-and 4-year-olds. This design reflects the hypothesis that different program impacts may be associated with different age of entry into Head Start. Differential impacts are of particular interest in light of a trend of increased enrollment of the 3-year-olds in some grantee/delegate agencies presumably due to the growing availability of preschool options for 4-year-olds. Consequently, the study included two separate samples: a newly entering 3-year-old group[3] (to be studied through two years of possible Head Start participation, kindergarten 1st grade, and 3rd grade), and a newly entering 4-year-old group (to be studied through one year of Head Start participation, kindergarten, 1st grade, and 3rd grade).

The study showed that the two age cohorts varied in demographic characteristics. The racial/ethnic characteristics of newly entering children in the 3-year-old cohort were substantially different from the characteristics of children in the newly entering 4-year-old cohort. While the newly entering 3-year-olds were relatively evenly distributed between Black children and Hispanic children about half of newly entering 4-year-olds were Hispanic children (see Exhibit 1). The ethnic difference was also reflected in the age-group differences in child and parent language.

Exhibit 1. Percentage of Children by Racial/Ethnic Characteristics and By Age Cohort

Child Race/Ethnicity	3-Year-Old Cohort	4-Year-Old Cohort
Hispanic	37.4%	51.6%
Black	32.8%	17.5%
White/Other	29.8%	30.8%

This study is unique in its design and differs from prior evaluations of early childhood programs:

- **Randomized Control.** The Congressional mandate for this study had a clearly stated goal of producing causal findings, i.e., the purpose was to determine if access to Head Start caused better developmental and parenting outcomes for participating children and families. To do this, the study randomly assigned Head Start applicants either to a Head Start group that was allowed to enroll, or to a "control" group that could not. This procedure ensured comparability between the two groups at program entry, so that later differences can be causally attributed to Head Start.
- **Representative Sample of Programs and Children.** Most random assignment studies are conducted in small demonstration programs or in a small number of operating sites, usually those that volunteer to be included in the research. In

[3] The study design allowed 3-year-old cohort control group children to reapply to Head Start after the first year.

contrast, the Head Start Impact Study is based on a nationally representative sample of Head Start programs and children, with a few exceptions for programs serving particular populations. This makes the study results generalizable to the vast majority of programs nationwide at the time the study was fielded in 2002, not just the selected study sample. Unlike most studies, it examines the average impact of programs that represent the full range of intensity and quality and adherence to the established Head Start program standards (i.e., the best, the worst, and those in the middle of a fully implemented program).

- **Examination of a Comprehensive Set of Outcomes Over Time.** The study quantifies the overall impact of Head Start separately for 3- and 4-year-old children in four key program domains-cognitive development, social-emotional development, health status and services, and parenting practices–following them through early elementary school. These impacts are measured by examining the difference in outcomes between children assigned to the Head Start group and those assigned to the control group.

Other study features that must be considered in interpreting the study findings include:

- **Control Group Children Did Not All Stay at Home.** Children who were placed in the control or comparison group were allowed to enroll in other non-parental care or non-Head Start child care or programs selected by their parents. They could remain at home in parent care, or enroll in a child care or preschool program. Consequently, the impact of Head Start was determined by a comparison to a mixture of alternative care settings rather than against a situation in which children were artificially prevented from obtaining child care or early education programs outside of their home. Approximately 60 percent of the control group children participated in child care or early education programs during the first year of the study, with 13.8 percent of the 4-year-olds in the control group and 17.8 percent of the 3-year-olds in the control group finding their way into Head Start during this year. Preventing families from seeking out alternative care or programs for their children is both infeasible and unethical. The design used here answers the policy question, how well does Head Start do when compared against the other types of services or care that low-income children could receive in fall 2002.

- **Impacts Represent the Effects of One Year of Head Start.** For children in the 4-year-old cohort, the study provides the impact of Head Start for a single year, i.e., the year before they are eligible to enter kindergarten. The impacts for the 3-year-old cohort reflect the benefits of being provided an earlier year of Head Start (as compared to the control group, which received access to Head Start at age 4.) At the end of one year of Head Start participation, the 3-year-old cohort—but not the 4-year-old cohort—had another year to go before they started kindergarten. It was not feasible or desirable for this study to prevent 3-year-olds from participating in Head Start for two years. Thus, the study could not directly assess the receipt of one year versus two years of Head Start. Rather, it addresses the receipt of an earlier year—whether having Head Start available at age three is helpful to children brought to the program at that age, or whether those children would be just as well off, if the program did not enroll them until age four. This is not only important to individual families; it also answers an important policy question. To answer this question, the

best approach is to preclude program entry at age three while allowing it at age four and contrast outcomes after that point with statistically equivalent children never excluded from the program. By design, the study did not attempt to control children's experiences after their first Head Start year.

The Head Start Impact Study is a comprehensive, carefully designed study of a large-scale early childhood program that has existed for more than 40 years. It is designed to address the overall average impact of the Head Start program as it existed in 2002. The findings cannot be directly compared to more narrowly focused studies of other early childhood programs. The Advisory Committee on Head Start Research and Evaluation, which developed the blueprint for this study, recommended that "the research and findings should be used in combination with the rest of the Head Start research effort to improve the effectiveness of Head Start programs for children and families" (Advisory Committee on Head Start Research and Evaluation, 1999, p. 44). The Third Grade Follow-up to the Head Start Impact Study builds upon the existing randomized control design in the HSIS in order to determine the longer term impact of the Head Start program on the well-being of children and families through the end of 3rd grade.

Key Findings

Looking across the full study period, from the beginning of Head Start through 3rd grade, the evidence is clear that access to Head Start improved children's preschool outcomes across developmental domains, but had few impacts on children in kindergarten through 3rd grade. Providing access to Head Start was found to have a positive impact on the types and quality of preschool programs that children attended, with the study finding statistically significant differences between the Head Start group and the control group on every measure of children's preschool experiences in the first year of the study. In contrast, there was little evidence of systematic differences in children's elementary school experiences through 3rd grade, between children provided access to Head Start and their counterparts in the control group.

In terms of children's well-being, there is also clear evidence that access to Head Start had an impact on children's language and literacy development while children were in Head Start. These effects, albeit modest in magnitude, were found for both age cohorts during their first year of admission to the Head Start program. However, these early effects rapidly dissipated in elementary school, with only a single impact remaining at the end of 3rd grade for children in each age cohort.

With regard to children's social-emotional development, the results differed by age cohort and by the person describing the child's behavior. For children in the 4-year-old cohort, there were no observed impacts through the end of kindergarten but favorable impacts reported by parents and unfavorable impacts reported by teachers emerged at the end of 1st and 3rd grades. One unfavorable impact on the children's self-report emerged at the end of 3rd grade. In contrast to the 4-year-old cohort, for the 3-year-old cohort there were favorable impacts on parent-reported social emotional outcomes in the early years of the study that continued into early elementary school. However, there were no impacts on teacher-reported measures of social-emotional development for the 3-year-old cohort at any data collection point or on the children's self-reports in 3rd grade.

In the health domain, early favorable impacts were noted for both age cohorts, but by the end of 3rd grade, there were no remaining impacts for either age cohort. Finally, with regard to parenting practices, the impacts were concentrated in the younger cohort. For the 4-year-old cohort, there was one favorable impact across the years while there were several favorable impacts on parenting approaches and parent-child activities and interactions (all reported by parents) across the years for the 3-year-old cohort.

In summary, there were initial positive impacts from having access to Head Start, but by the end of 3rd grade there were very few impacts found for either cohort in any of the four domains of cognitive, social-emotional, health and parenting practices. The few impacts that were found did not show a clear pattern of favorable or unfavorable impacts for children.

In addition to looking at Head Start's average impact across the diverse set of children and families who participated in the program, the study also examined how impacts varied among different types of participants. There is evidence that for some outcomes, Head Start had a differential impact for some subgroups of children over others. At the end of 3rd grade for the 3-year-old cohort, the most striking sustained subgroup findings were found in the cognitive domain for children from high risk households as well as for children of parents who reported no depressive symptoms. Among the 4-year-olds, sustained benefits were experienced by children of parents who reported mild depressive symptoms, severe depressive symptoms, and Black children.

Overview of Study Methods

Random Assignment

Newly entering 3- and 4-year-old Head Start applicants were randomly assigned either to a Head Start group that for one year 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 the research questions outlined by Congress, 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 Head Start group that had access to Head Start services in the initial year of the study or to a control group that could receive any other non-Head Start services available in the community, chosen by their parents. In fact, approximately 60 percent of control group parents enrolled their children in some other type of preschool program in the first year. In addition, all children in the 3-year-old cohort could receive Head Start services in the second year. 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 in the initial year on children's school readiness. This research design ensured that the Head Start and control groups did not differ in any systematic or unmeasured way except through their access to Head Start services. It is important to note that, because the control group in the 3-year-old cohort was given access to Head Start in the second year, the findings for this age group reflect the added benefit of providing access to Head Start at age 3 vs. at age 4, *not* the total benefit of having access to Head Start for two years.

In addition to random assignment, this study is set apart from most program evaluations because it includes a nationally representative sample of programs, making results generalizable to the Head Start program as a whole, not just to the selected samples of programs

and children. However, the study does not represent Head Start programs serving special populations, such as tribal Head Start programs, programs serving migrant and seasonal farm workers and their families, or Early Head Start. Further, the study does not represent the 15 percent of Head Start programs in which the pool of applicants for Head Start slots was too small to allow for an adequate control group.

Study Sample

The nationally representative study sample, spread over 23 different states, consisted 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.

Selected Head Start grantees and centers had to have a sufficient number of applicants for the 2002-2003 program year to allow for the creation of a control group without requiring Head Start slots to go unfilled. 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 2002-2003 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 control group. Random assignment was done separately for two study samples—newly entering 3-year-olds (to be studied through two years of potential Head Start participation, kindergarten, 1st grade, and 3rd grade) and newly entering 4-year-olds (to be studied through one year of Head Start participation, kindergarten, 1st grade, and 3rd grade).

The total sample, spread over 23 different states, consisted 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.[4]

Data collection began in the fall of 2002 and continued through the spring of 2008, following children from entry into Head Start through the end of 3rd grade. Comparable data were collected for both Head Start and control group children, including interviews with parents, direct child assessments, surveys of Head Start, other early childhood, and elementary school teachers, interviews with center directors and other care providers at the preschool level, direct observations of the quality of various preschool care settings, and teacher or care provider assessments of children. For the Third Grade Follow-up, principal surveys and teacher ratings by the principal were added to the data collection. Response rates were consistently quite high, approximately 80 percent for parents and children throughout the study. Teacher response rates were higher at the preschool level (about 80 percent) and gradually decreased as the child

[4] The sample of 3-year-olds is slightly larger than the sample of 4-year-olds to ensure that an adequate sample size was maintained, given the possibility of higher study attrition resulting from an additional year of longitudinal data collection for the younger children.

progressed through school (slightly above 60 percent by the end of 3rd grade). Principal data were collected only during 3rd grade and the response rate was about the same as for 3rd grade teachers.

Although every effort was made to ensure compliance with random assignment, some children accepted into Head Start did not participate in the program (about 15 percent for the 3-year-old cohort and 20 percent for the 4-year-old cohort), and some children assigned to the non-Head Start group nevertheless entered the program in the first year (about 17 percent for 3-year-olds and 14 percent for 4-year-olds), typically at centers that were not in the study sample. These families are referred to as "no shows" and "crossovers." Statistical procedures for dealing with these events are discussed in the report. Thus, the findings in this report provide estimates of both the impact of *access* to Head Start using the sample of all randomly assigned children (referred to as Intention to Treat, or ITT) and the impact of *actual* Head Start participation (adjusting for the no shows and crossovers, referred to as Impacts on the Treated or IOT).

Findings: Head Start Through 3rd Grade

Impact on Children's Experiences in Preschool and Early Elementary School

There are clear impacts on the types and quality of children's child care, early education, and school experiences at the preschool level but not in the early elementary grades. At the preschool level, the story is far clearer, as providing access to Head Start was found to have a positive impact on children's experiences across many measures of early childhood experience. There were statistically significant differences between the Head Start group and the control group on every measure of children's preschool experiences measured in this study. These effects were found both for the 4-year-old cohort and for the 3-year-old cohort during the year in which they were admitted to Head Start. The measures that were examined included, but were not limited to, teacher qualifications, including their training and education; classroom literacy and math instructional activities; classroom teacher-child ratios; the nature of teacher-child interactions; and global measures of the care environment as measured by ECERS-R/ FDCRS scores. The differences in magnitude were quite large, driven in part by the large proportion of children in the control group who were in parent care (i.e., nearly four out of ten children remained at home with their parents when Head Start was unavailable to them).[5] Yet, analyses excluding those children, and thus comparing only children in the Head Start and control groups who were in non-parental care, largely showed the same pattern of differences, albeit somewhat smaller.

The preschool experiences of children in the 3-year-old cohort were very different in the second year of the study, when most were 4 years old. The majority of the children (both Head Start and control group) were in some type of center-based care by the this year, and with three small exceptions, the observed treatment and control differences disappeared in the age 4 year. That is, once the control group had access to Head Start, the earlier differences on the measures of their early childhood care environments all but vanished.

[5] For these analyses, children in parent care were included and given a score of zero.

Providing access to Head Start did not have much impact on the types of schools children attended from kindergarten through 3rd grade. By the end of 3rd grade, the study sample had dispersed into nearly 1,600 individual elementary schools. On average, both Head Start and control group children attended public schools, with the percentage enrolled in public school increasing from kindergarten to 3rd grade for children in the study sample as a whole. For the 4-year-old cohort, approximately 80 percent of the children were enrolled in public school at the end of kindergarten. By the end of 3rd grade, 98 percent of the children were enrolled in public school. For the 3-year-old cohort, approximately 85 percent were enrolled in public schools at the end of kindergarten and 96 percent by the end of the 3rd grade. The schools' percentages of students at or above proficient on state assessments in math and reading were in the middle of the respective state averages (55 to 67 percent depending on the subject and year), indicating that on average the schools attended by the study children were not among the worst or best schools in their respective states. In the 3-year-old cohort's kindergarten year, a significant difference was found in the school-wide average math proficiency scores for Head Start children and control group children, with the difference favoring the control group. For the 3-year-old cohort in the 3rd grade, a significant difference was found between average reading/language arts proficiency scores at the schools attended by the Head Start and control group children, this time favoring the Head Start group.

Not surprisingly, the study children attended schools with much higher levels of poverty than schools nationwide (as indicated by proportions of students eligible for free- and reduced-price lunch—66-67 percent) and were in schools with higher proportions of minority students (approximately 60 percent of students). With only a few exceptions, teacher and classroom characteristics did not differ significantly between children in the Head Start group and those in the control group. The school experiences measures were limited in kindergarten and 1st grade, while a wider range of school, classroom, and teacher measures assessed the children's 3rd grade year. The few differences that were found varied and most were found at the end of 3rd grade, sometimes favoring the control group and sometimes favoring the Head Start group.

For children in the 4-year-old cohort at the end of 3rd grade year, there is evidence that the Head Start children were in schools that, according to principals, had greater access to computers (compared to the non-Head Start children), and were more likely to have summer school programs. Head Start children were more likely to have a teacher with a standard state teaching certificate, and their schools were more likely to use standardized tests to a great extent to compare subgroups of students. On the other hand, the schools attended by the Head Start children were more likely to have higher levels of student mobility.

At the end of 3rd grade for children in the 3-year-old cohort, principal reports showed that Head Start children were in schools that had more adequate school facilities, lower staff turnover, and a higher percentage of 3rd grade students scoring at the proficient or higher level on the state reading/language arts assessment. On the other hand, the schools attended by Head Start children had higher percentages of students with disabilities, and according to principals, required more attention to deal with student discipline problems. In terms of classroom and teacher measures, the Head Start children were more likely to have access to an instructional aide in their classroom, to have a lower percentage of children reading below grade level in their reading/language arts class, and to have a teacher who majored in education as an undergraduate.

Impacts on Children's Cognitive Development

The cognitive domain consisted of: (1) direct assessments of language and literacy skills, pre-writing skills (in Head Start years only), and math skills; (2) teacher reports of children's school performance; and (3) parent reports of child literacy skills and grade promotion.

There is clear evidence that Head Start had a statistically significant impact on children's language and literacy development while children were in Head Start. These effects, albeit modest in magnitude, were found for both age cohorts during their first year of admission to the Head Start program. However, these early effects dissipated in elementary school, with only a single impact remaining at the end of 3rd grade for children in each age cohort: a favorable impact for the 4-year-old cohort (ECLS-K Reading) and an unfavorable impact for the 3-year-old cohort (grade promotion).

Impacts aside, these children remain disadvantaged compared to their same-age peers; the scores of both the Head Start and the control group children remained lower than the norm for the population. At the end of 3rd grade, HSIS children (both Head Start and control group children) in the 4-year-old cohort, on average, scored about eight points (approximately one-half of a standard deviation) lower than a national sample of third graders on the ECLS-K Reading Assessment and the promotion rate[6] for the 3-year old cohort was two to three percent lower than the predicted national promotion rate for children at the end of 3rd grade.

For mathematics**,** impacts were found only on a single outcome measure (Woodcock Johnson III Applied Problems) and only for the 3-year-old cohort at the end of their Head Start year.

The findings from the cognitive domain are summarized by age cohort below.[7] Exhibits 2a and 2b present all statistically significant cognitive impacts and their effect sizes[8] from the Intent to Treat (ITT) analysis.

4-Year-Old Cohort

- At the end of the Head Start year, there was strong evidence that the Head Start group demonstrated better skills on the following six child outcomes related to children's language and literacy development: (1) Peabody Picture Vocabulary Test (PPVT) (vocabulary); (2) Woodcock-Johnson III (WJIII) Letter-Word Identification; (3) WJIII Spelling; (4) WJIII Pre-Academic Skills; (5) Color Identification; and (6) Letter Naming.

---

[6] Warren and Saliba (2012) generated a predicted 3rd grade national retention rate using an age-grade delay model as a proxy for retention. See Chapter 4 for additional information.

[7] Three levels of evidence are considered in this report: (1) strong evidence is used for impacts statistically significant at $p \leq 0.05$, and the result holds up under the Benjamini-Hochberg test for multiple comparisons; (2) moderate evidence signifies a particular impact is statistically significant at $p \leq 0.05$, but this result does not hold up under the test for multiple comparisons; and (3) suggestive evidence signifies a particular impact is statistically significant under a relaxed standard $p \leq 0.10$, and the result may or may not hold up under the test for multiple comparisons.

[8] The effect size is simply the impact estimate divided by the standard deviation of the outcome measure in the population. The effect size provides an indication of the magnitude of each impact that is independent of the particular instrument or measure used. More discussion of the interpretation of effect sizes is provided in Chapter 2.

- Parents of children in the Head Start group reported that their children had greater emerging literacy skills at the end of Head Start than did parents of children in the control group.
- There were no impacts for 4-year-olds in the cognitive domain at the end of kindergarten.
- At the end of 1st grade, there was suggestive evidence of a positive impact of access to Head Start on PPVT (vocabulary) scores.
- At the end of 3rd grade, there was suggestive evidence of a positive impact of access to Head Start on the ECLS-K Reading Assessment.
- No significant impacts were found for math skills, pre-writing, children's promotion, or teacher report of children's school accomplishments or abilities in any year.

3-Year-Old Cohort

- At the end of their Head Start year, there was strong evidence of better skills for the Head Start group on the following five outcomes related to children's language and literacy development: (1) PPVT (vocabulary), (2) WJIII Letter-Word, (3) Preschool Comprehensive Test of Phonological and Print Processing (CTOPPP) Elision, (4) Letter Naming, and (5) WJIII Pre-Academic Skills. There was also a statistically significant impact on the measure of children's pre-writing skills. Children in the Head Start group were found to have more advanced math skills than their counterparts at the end of the Head Start year on the WJIII test of Applied Problems.
- Favorable impacts of Head Start were also found on parental reports of children's emergent literacy skills at the end of the Head Start year.
- At the end of the age 4 year, few statistically significant impacts were found. However, two impacts persisted related to children's literacy skills. Children in the Head Start group scored higher than children in the control group on CTOPPP Elision as well as on parents' reports of their literacy skills.
- As with the 4-year-old cohort, there was no strong evidence of impacts on children's language, literacy, or math measures at the end of kindergarten or at the end of 1st grade. However, there was suggestive evidence of an impact on Oral Comprehension at the end of 1st grade.
- At the end of 3rd grade, there was suggestive evidence of an unfavorable impact—the parents of the Head Start group children reported a significantly lower child grade promotion rate than the parents of the non-Head Start group children.
- No statistically significant impacts were found for teacher reports of children's school performance, with the exception of a lower teacher assessment in kindergarten of Head Start children's math ability. This was not supported by children's scores on the three direct math assessments, which showed no evidence of math differences. However, the schools attended by the control group children in the 3-year-old cohort during their kindergarten year reported a significantly higher percentage of students at or above the proficient level in math than the schools attended by the Head Start group children.

Exhibit 2a. Summary of ITT Cognitive Impacts for 4-Year-Olds by Year

Measure	Age 4 (Head Start Year)	K	1st Grade	3rd Grade
Language, Literacy, and Pre-Writing				
Color Identification	**0.16**	NA	NA	NA
Pre-Writing (McCarthy Draw a Design)		NA	NA	NA
Emergent Literacy Scale (parent report)	**0.31**	NA	NA	NA
Letter Naming	**0.25**		NA	NA
Test of Phonological Processing (CTOPPP Elision)			NA	NA
Receptive Vocabulary (PPVT)	**0.09**		**0.09**	
Letter-Word Identification (WJIII)	**0.22**			
Spelling (WJIII)	**0.15**			NA
Oral Comprehension (WJIII)				NA
Pre-Academic Skills (WJIII)	**0.19**			NA
Phonetic Skills/ Word Attack (WJIII)	NA			NA
Basic Reading (WJIII)	NA			NA
Academic Applications (WJIII)	NA	NA		NA
Academic Skills (WJIII)	NA	NA		NA
Passage Comprehension (WJIII)	NA	NA		NA
ECLS-K Reading	NA	NA		**0.11**
Writing Sample (WJIII)	NA	NA		NA
Spanish Language				
Receptive Vocabulary (TVIP)				NA
Batería WM Identificación de letras y palabras				
Math				
One-to-One Counting (Counting Bears)		NA	NA	NA
Applied Problems (WJIII)				
Quantitative Concepts (WJIII)	NA			NA
Math Reasoning (WJIII)	NA			NA
Calculation (WJIII)	NA	NA		
School Performance				
School Accomplishments	NA			NA
Promotion (parent report)	NA			
Language and Literacy Ability	NA			
Math Ability	NA			
Math Skills	NA	NA	NA	
Reading/Language Arts Skills	NA	NA	NA	
Social Studies and Science Ability	NA			NA

KEY:

Dark gray cell indicates a significant favorable impact ($p \leq 0.10$).

Diagonal mark cell indicates a significant unfavorable impact ($p \leq 0.10$).

Blank cell indicates no significant impact.

An * indicates that the interpretation of the outcome is unclear.

NA indicates that no data were collected for this outcome at this data collection point.

NOTE: Intent to Treat (ITT) effect sizes are presented only for statistically significant differences ($p \leq 0.10$). The effect size is simply the impact estimate divided by the standard deviation of the outcome measure in the population. The effect size provides an indication of the magnitude of each impact that is independent of the particular instrument or measure used. More discussion of the interpretation of effect sizes is provided in Chapter 2.

Exhibit 2b. Summary of ITT Cognitive Impacts for 3-Year-Olds by Year

Measure	Age 3 (Head Start Year)	Age 4	K	1st Grade	3rd Grade
Language, Literacy, and Pre-Writing					
Color Identification			NA	NA	NA
Pre-Writing (McCarthy Draw a Design)	**0.14**		NA	NA	NA
Emergent Literacy Scale (parent report)	**0.35**	**0.16**	NA	NA	NA
Letter Naming	**0.24**			NA	NA
Test of Phonological Processing (CTOPPP Elision)	**0.10**	**0.15**		NA	NA
Receptive Vocabulary (PPVT)	**0.18**				
Letter-Word Identification (WJIII)	**0.26**				
Spelling (WJIII)					NA
Oral Comprehension (WJIII)				**0.08**	NA
Pre-Academic Skills (WJIII)	**0.22**				NA
Phonetic Skills/Word Attack (WJIII)	NA	NA			NA
Basic Reading (WJIII)	NA	NA			NA
Academic Applications (WJIII)	NA	NA	NA		NA
Academic Skills (WJIII)	NA	NA	NA		NA
Passage Comprehension (WJIII)	NA	NA	NA		NA
ECLS-K Reading	NA	NA	NA		
Writing Sample (WJIII)	NA	NA	NA		NA
Spanish Language					
Receptive Vocabulary (TVIP)					NA
Batería WM Identificación de letras y palabras			**0.26**		
Math					
One-to-One Counting/Counting Bears			NA	NA	NA
Applied Problems (WJIII)	**0.15**				
Quantitative Concepts (WJIII)	NA	NA			NA
Math Reasoning (WJIII)	NA	NA			NA
Calculation (WJIII)	NA	NA	NA		
School Performance					
School Accomplishments	NA	NA			NA
Promotion (parent report)	NA	NA			**-0.11**
Language and Literacy Ability	NA	NA			
Math Ability	NA	NA	**-0.19**		
Math Skills	NA	NA	NA	NA	
Reading/Language Arts Skills	NA	NA	NA	NA	
Social Studies and Science Ability	NA	NA			NA

KEY:

Dark gray cell indicates a significant favorable impact ($p \leq 0.10$).

Diagonal mark cell indicates a significant unfavorable impact ($p \leq 0.10$).

Blank cell indicates no significant impact.

An * indicates that the interpretation of the outcome is unclear.

NA indicates that no data were collected for this outcome at this data collection point.

NOTE: Intent to Treat (ITT) effect sizes are presented only for statistically significant differences ($p \leq 0.10$). The effect size is simply the impact estimate divided by the standard deviation of the outcome measure in the population. The effect size provides an indication of the magnitude of each impact that is independent of the particular instrument or measure used. More discussion of the interpretation of effect sizes is provided in Chapter 2.

Impacts on Children's Social-Emotional Development

The social-emotional domain consisted of parent-reported measures during the Head Start years, reports by both parents and teachers in all elementary school years, with child self-reports added at the end of 3rd grade. Measures of children's behavior, social skills and approaches to learning, parent-child relationships, teacher child relationships, school adjustment, peer relationships and school experiences were assessed.

With regard to children's social-emotional development, the results differed by age cohort and by the source of the information on the child's behavior. For children in the 4-year-old cohort, there were no observed impacts through the end of kindergarten and then favorable impacts reported by parents and unfavorable impacts reported by teachers at the end of 1st and 3rd grades and children at the end of 3rd grade.

In contrast, the early favorable social emotional impacts reported by parents for the 3-year-old cohort continued into early elementary school. There were favorable impacts at all data collection points through the end of 3rd grade on parent-reported measures of children's social-emotional development. However, there were no impacts on teacher-reported measures of social-emotional development for the 3-year-old cohort at any data collection point or on the children's self-reports in 3rd grade.

The findings from the social-emotional domain are summarized by age cohort below. Exhibits 3a and 3b provide all statistically significant social-emotional impacts and their effect sizes from the ITT analysis.

4-Year-Old Cohort

- There were no significant differences between the Head Start group and the control group on any measures of social-emotional development during the Head Start year or during kindergarten.
- At the end of 1st grade, impacts on social-emotional development were few and mixed.
  - There were two unfavorable findings based on teacher reports of children's behavior: (1) children in the Head Start group demonstrated moderate evidence of more socially reticent behavior (i.e., shy and hesitant behavior) as reported by teachers, and (2) there was suggestive evidence of more problematic student-teacher interactions.
  - In contrast, there was suggestive evidence of less withdrawn behavior for children in the Head Start group as reported by their parents.
- At the end of 3rd grade, parents reported less aggressive and total problem behaviors for the Head Start group children. However, teachers reported unfavorable impacts with a higher incidence of children's emotional symptoms, less closeness, and a less positive relationship with the Head Start children. Finally, Head Start children in the 4-year-old cohort reported less positive peer relations at school compared to the control group.

3-Year-Old Cohort

- At the end of the Head Start year, children in the Head Start group showed strong evidence of less hyperactive behavior and fewer overall problem behaviors as reported by their parents.
- At the end of the age 4 year and the end of kindergarten, children in the Head Start group demonstrated suggestive evidence of better social skills and positive approaches to learning as reported by their parents. Further, children in the Head Start group also continued to show moderate evidence of less hyperactive behavior at the end of kindergarten.
- By the end of 1st grade, parents of Head Start group children reported moderate evidence of a closer relationship with their child than parents of control group children. At the same time, parents of Head Start group children reported (suggestive evidence) a more positive overall relationship with their child than parents of children in the control group.
- There were no impacts on teacher-reported measures of social-emotional development for the 3-year-old cohort in either the kindergarten or 1st grade year.
- For this age cohort, there was only a single statistically significant social-emotional impact at the end of 3rd grade. Children in the Head Start group demonstrated better social skills and positive approaches to learning as reported by their parents, compared with the non-Head Start group.

Exhibit 3a. Summary of ITT Social-Emotional Impacts for 4-Year-Olds by Year

Measure	Age 4 (Head Start Year)	K	1st Grade	3rd Grade
Parent-Reported Measures				
Aggressive Behavior				**-0.13**
Hyperactive Behavior				
Withdrawn Behavior			**-0.13**	
Total Problem Behavior				**-0.12**
Social Competencies				NA
Social Skills and Positive Approaches To Learning				
Closeness with Parent				NA
Conflict with Parent				NA
Positive Parent-Child Relationships				NA
Teacher-Reported Measures				
Aggressive (ASPI)	NA			NA
Interactive/Hyperactive (ASPI)	NA			
Withdrawn/Low Energy (ASPI)	NA			NA
Oppositional (ASPI)	NA			NA
Problems with Peer Interaction (ASPI)	NA			NA
Shy/Socially Reticent (ASPI)	NA		**0.19**	NA
Problems with Structured Learning (ASPI)	NA			NA
Problems with Teacher Interaction (ASPI)	NA		**0.13**	NA
Closeness with Teacher	NA			**-0.13**
Conflict with Teacher	NA			
Positive Teacher-Child Relationships	NA			**-0.14**
Conduct Problems-% in Normal Category	NA	NA	NA	
Emotional Symptoms-% in Normal Category	NA	NA	NA	**-0.24**
Hyperactivity-% in Normal Category	NA	NA	NA	
Peer Problems-% in Normal Category	NA	NA	NA	
Pro-social Behavior-% in Normal Category	NA	NA	NA	
Total Difficulties-% in Normal Category	NA	NA	NA	
Social Competency	NA	NA	NA	
Child-Reported Measures				
Externalizing	NA	NA	NA	
Internalizing	NA	NA	NA	
Peer Relations	NA	NA	NA	**-0.14**
School	NA	NA	NA	

KEY:

Dark gray cell indicates a significant favorable impact ($p \leq 0.10$).

Diagonal mark cell indicates a significant unfavorable impact ($p \leq 0.10$).

Blank cell indicates no significant impact.

An * indicates that the interpretation of the outcome is unclear.

NA indicates that no data were collected for this outcome at this data collection point.

NOTE: Intent to Treat (ITT) effect sizes are presented only for statistically significant differences ($p \leq 0.10$). The effect size is simply the impact estimate divided by the standard deviation of the outcome measure in the population. The effect size provides an indication of the magnitude of each impact that is independent of the particular instrument or measure used. More discussion of the interpretation of effect sizes is provided in Chapter 2.

Exhibit 3b. Summary of ITT Social-Emotional Impacts for 3-Year-Olds by Year

Measure	Age 3 (Head Start Year)	Age 4	K	1st Grade	3rd Grade
Parent-Reported Measures					
Aggressive Behavior					
Hyperactive Behavior	**-0.21**		**-0.12**		
Withdrawn Behavior					
Total Problem Behavior	**-0.14**				
Social Competencies					**NA**
Social Skills and Positive Approaches To Learning		**0.11**	**0.14**		**0.12**
Closeness with Parent				**0.10**	NA
Conflict with Parent					NA
Positive Parent-Child Relationships				**0.10**	NA
Teacher-Reported Measures					
Aggressive (ASPI)	NA	NA			**NA**
Interactive/Hyperactive (ASPI)	NA	NA			
Withdrawn/Low Energy (ASPI)	NA	NA			NA
Oppositional (ASPI)	NA	NA			NA
Problems with Peer Interaction (ASPI)	NA	NA			NA
Shy/Socially Reticent (ASPI)	NA	NA			NA
Problems with Structured Learning (ASPI)	NA	NA			NA
Problems with Teacher Interaction (ASPI)	NA	NA			NA
Closeness with Teacher	NA	NA			
Conflict with Teacher	NA	NA			
Positive Teacher-Child Relationships	NA	NA			
Conduct Problems-% in Normal Category	NA	NA	NA	NA	
Emotional Symptoms-% in Normal Category	NA	NA	NA	NA	
Hyperactivity-% in Normal Category	NA	NA	NA	NA	
Peer Problems-% in Normal Category	NA	NA	NA	NA	
Pro-social Behavior-% in Normal Category	NA	NA	NA	NA	
Total Difficulties-% in Normal Category	NA	NA	NA	NA	
Social Competency	NA	NA	NA	NA	
Child-Reported Measures					
Externalizing	NA	NA	NA	NA	
Internalizing	NA	NA	NA	NA	
Peer Relations	NA	NA	NA	NA	
School	NA	NA	NA	NA	

KEY:

Dark gray cell indicates a significant favorable impact ($p \leq 0.10$).

Diagonal mark cell indicates a significant unfavorable impact ($p \leq 0.10$).

Blank cell indicates no significant impact.

An * indicates that the interpretation of the outcome is unclear.

NA indicates that no data were collected for this outcome at this data collection point.

NOTE: Intent to Treat (ITT) effect sizes are presented only for statistically significant differences ($p \leq 0.10$). The effect size is simply the impact estimate divided by the standard deviation of the outcome measure in the population. The effect size provides an indication of the magnitude of each impact that is independent of the particular instrument or measure used. More discussion of the interpretation of effect sizes is provided in Chapter 2.

Impact on Health Status and Access to Health Services

The health domain consisted of two categories: (1) children's receipt of health care services and (2) their current health status. Early favorable impacts in the health domain were noted for both age cohorts but by the end of 3rd grade, there were no remaining impacts for either age cohort.

The findings from the health domain are summarized by age cohort below, while Exhibits 4a and 4b present all statistically significant health impacts and their effect sizes from the ITT analysis.

4-Year-Old Cohort

- At the end of the Head Start year, there was strong evidence that access to Head Start increased children's receipt of dental care—a difference of 15 percentage points.
- In kindergarten, there was suggestive evidence of an improvement in children's health status and an increase in health insurance coverage (differences of five and four percentage points, respectively).
- By the end of 1st grade, there was still moderate evidence of increased health insurance coverage among the Head Start group —a difference of four percentage points.
- There were no significant impacts at the end of 3rd grade.

3-Year-Old Cohort

- At the end of the Head Start year and again at the end of the age 4 year, there was strong evidence that access to Head Start increased children's receipt of dental care—differences of 17 and 10 percentage points, respectively.
- There was moderate evidence of improvements on children's reported overall health status at the end of the Head Start year and moderate evidence of an impact on health insurance coverage at the end of kindergarten.
- There was evidence of a significant impact on care for injuries[9] at the end of the age 4 year, although the interpretation of this impact is unclear.
- There were no significant impacts at the end of 1st or 3rd grades.

[9] The interpretation of child had care for injury in the last month is unclear. The change may reflect an increase in injuries, an increase in care-seeking, or both.

Exhibit 4a. Summary of ITT Health Impacts for 4-Year-Olds by Year

Measure	Age 4 (Head Start Year)	K	1st Grade	3rd Grade
Parent-Reported Measures				
Child Received Dental Care	**0.31**			
Child Has Health Insurance Coverage		**0.11**	**0.11**	
Child's Overall Health Status is Excellent/ Good		**0.13**		
Child Needs Ongoing Health Care				
Child Had Care for Injury in Last Month				

KEY:

Dark gray cell indicates a significant favorable impact ($p \leq 0.10$).

Diagonal mark cell indicates a significant unfavorable impact ($p \leq 0.10$).

Blank cell indicates no significant impact.

An * indicates that the interpretation of the outcome is unclear.

NA indicates that no data were collected for this outcome at this data collection point.

NOTE: Intent to Treat (ITT) effect sizes are presented only for statistically significant differences ($p \leq 0.10$). The effect size is simply the impact estimate divided by the standard deviation of the outcome measure in the population. The effect size provides an indication of the magnitude of each impact that is independent of the particular instrument or measure used. More discussion of the interpretation of effect sizes is provided in Chapter 2.

Exhibit 4b. Summary of ITT Health Impacts for 3-Year-Olds by Year

Measure	Age 3 (Head Start Year)	Age 4	K	1st Grade	3rd Grade
Parent-Reported Measures					
Child Received Dental Care	**0.33**	**0.20**			
Child Has Health Insurance Coverage			**0.14**		
Child's Overall Health Status is Excellent/ Good	**0.11**				
Child Needs Ongoing Health Care					
Child Had Care for Injury in Last Month[10]		**0.10***			

KEY:

Dark gray cell indicates a significant favorable impact ($p \leq 0.10$).

Diagonal mark cell indicates a significant unfavorable impact ($p \leq 0.10$).

Blank cell indicates no significant impact.

An * indicates that the interpretation of the outcome is unclear.

NA indicates that no data were collected for this outcome at this data collection point.

NOTE: Intent to Treat (ITT) effect sizes are presented only for statistically significant differences ($p \leq 0.10$). The effect size is simply the impact estimate divided by the standard deviation of the outcome measure in the population. The effect size provides an indication of the magnitude of each impact that is independent of the particular instrument or measure used. More discussion of the interpretation of effect sizes is provided in Chapter 2.

[10] See footnote 9.

Impact on Parenting Practices

This domain consisted of six categories of outcomes: (1) disciplinary practices, (2) educational supports, (3) safety practices, (4) parenting styles, (5) parent participation in and communication with school and (6) parent and child time together. With regard to parenting practices, the impacts were concentrated in the younger cohort, which showed favorable parent-reported impacts across all years of the study. For the 4-year-old cohort, in contrast, there were few impacts.

The findings from the parenting practices domain are summarized by age cohort below, and Exhibits 5a and 5b provide the statistically significant parenting practices impacts and their effect sizes from the ITT analysis.

4-Year-Old Cohort

- There were minimal impacts for the 4-year-old cohort in this domain, with two exceptions: at the end of the Head Start year, parents in the Head Start group were less likely to use time out[11] as a disciplinary practice than were parents in the control group and at the end of 3rd grade, there was strong evidence of a large favorable impact on parental reports of the amount of time they spent with their child.

3-Year-Old Cohort

- In the Head Start year, there were three impacts on parenting practices, of which two impacts (spanking and cultural enrichment) were supported by strong evidence:
  - Parents of children in the Head Start group were less likely to have spanked their children than parents in the control group (a difference of seven percentage points).
  - Parents of children in the Head Start group were more likely to have read to their child in the last week than parents in the control group.
  - Parents of children in the Head Start group involved their child in cultural enrichment activities more than parents of children in the control group.
- At the end of the age 4 year, there was a favorable impact on parenting, with parents of children in the Head Start group less likely to use an authoritarian parenting style (characterized by high control and low warmth) than parents of children in the control group.
- Evidence of impacts on parenting continued in kindergarten, 1st and 3rd grades.
  - At the end of kindergarten, there was suggestive evidence that parents of children in the Head Start group were less likely to spank their children and moderate evidence that these parents were less likely to use time out.[12]

[11] The interpretation of time out is unclear. The change may reflect favorable changes in the children's behavior, changes in the parents' reactions (whether to less or more desirable forms of discipline), or both.

[12] See footnote 11.

- At the end of 1st grade, there was also suggestive evidence that parents of children in the Head Start group were less likely to use time out and moderate evidence that these parents were less likely to use an authoritarian parenting style.
- At the end of 3rd grade, there was a favorable impact on the use of the preferred authoritative parenting style (characterized by high warmth and high control).

Exhibit 5a. Summary of ITT Parenting Practices Impacts for 4-Year-Olds by Year

Measure	Age 4 (Head Start Year)	K	1st Grade	3rd Grade
Parent-Reported Measures				
Parent Spanked Child in Last Week				
Parent Used Time Out in Last Week[13]	**-0.17***			NA
Parent Read to Child in Last Week				NA
Parental Safety Practices Scale				NA
Family Cultural Enrichment Scale				NA
Parenting Style: Authoritarian	NA			
Parenting Style: Authoritative	NA			
Parenting Style: Neglectful	NA			
Parenting Style: Permissive	NA			
Supportive School Environment	NA	NA	NA	
Effect of Parenting on Parent's Life	NA	NA	NA	
Doing Things Together	NA	NA	NA	
Time Spent with Child	NA	NA	NA	**0.27**
Parent Perception of School Services	NA	NA	NA	
Teacher-Reported Measures				
School Contact and Communication	NA			
Parent Participation	NA			

KEY:

Dark gray cell indicates a significant favorable impact ($p \leq 0.10$).

Diagonal mark cell indicates a significant unfavorable impact ($p \leq 0.10$).

Blank cell indicates no significant impact.

An * indicates that the interpretation of the outcome is unclear.

NA indicates that no data were collected for this outcome at this data collection point.

NOTE: Intent to Treat (ITT) effect sizes are presented only for statistically significant differences ($p \leq 0.10$). The effect size is simply the impact estimate divided by the standard deviation of the outcome measure in the population. The effect size provides an indication of the magnitude of each impact that is independent of the particular instrument or measure used. More discussion of the interpretation of effect sizes is provided in Chapter 2.

[13] See footnote 11.

Exhibit 5b. Summary of ITT Parenting Practices Impacts for 3-Year-Olds by Year

Measure	Age 3 (Head Start Year)	Age 4	K	1st Grade	3rd Grade
Parent-Reported Measures					
Parent Spanked Child in Last Week	**-0.14**		**-0.09**		
Parent Used Time Out in Last Week[14]			**-0.13***	**-0.11***	NA
Parent Read to Child in Last Week	**0.15**				NA
Parental Safety Practices Scale					NA
Family Cultural Enrichment Scale	**0.18**				NA
Parenting Style: Authoritarian	NA	**-0.14**		**-0.11**	
Parenting Style: Authoritative	NA				**0.16**
Parenting Style: Neglectful	NA				
Parenting Style: Permissive	NA				
Supportive School Environment	NA	NA	NA	NA	
Effect of Parenting on Parent's Life	NA	NA	NA	NA	
Doing Things Together	NA	NA	NA	NA	
Time Spent with Child	NA	NA	NA	NA	
Parent Perception of School Services	NA	NA	NA	NA	
Teacher-Reported Measures					
School Contact and Communication	NA	NA			
Parent Participation	NA	NA			

KEY:

Dark gray cell indicates a significant favorable impact ($p \leq 0.10$).

Diagonal mark cell indicates a significant unfavorable impact ($p \leq 0.10$).

Blank cell indicates no significant impact.

An * indicates that the interpretation of the outcome is unclear.

NA indicates that no data were collected for this outcome at this data collection point.

NOTE: Intent to Treat (ITT) effect sizes are presented only for statistically significant differences ($p \leq 0.10$). The effect size is simply the impact estimate divided by the standard deviation of the outcome measure in the population. The effect size provides an indication of the magnitude of each impact that is independent of the particular instrument or measure used. More discussion of the interpretation of effect sizes is provided in Chapter 2.

Variation in Impact

This report examines differences in impact among different groups of children and parents. Seven dimensions were used to define subgroups: (1) whether a child had low pre-academic skills, (2) whether the child was a Dual Language Learner, (3) whether the child had special needs (as reported by the parent), (4) mother's race/ethnicity, (5) reported level of depressive symptoms for the child's parent/primary caregiver, (6) a composite index of household risks, and (7) urban location. All categorizations were based on data collected at the time of random assignment. Sample sizes by subgroup, age cohort, and random assignment status are presented in Chapter 5.

[14] See footnote 11.

The approach to analyzing subgroups was to highlight patterns in the findings. There is no scientific consensus for what constitutes a pattern of impacts. Yet, given the large number of comparisons tested (over 13,000, taking into consideration the study's two cohorts, five time points for measuring outcomes, and multiple outcomes across many subgroups), it was important to find an approach that balances the risk of reporting on chance findings with that of ignoring important findings. To this end, the subgroup findings concentrate on differential impacts, that is, impacts where there was a statistically significant difference in Head Start's effects for one subgroup compared to another. Accordingly, the discussion primarily focuses on results where there was both a statistically significant difference in impacts between subgroups and a statistically significant impact for at least one subgroup in the comparison.

Particular attention was paid to end of 3rd grade results that showed a pattern across domains and how those results related to prior time points.[15] The subgroup findings should be viewed as secondary and exploratory as compared to the main impact findings that are considered primary as well as confirmatory.

At the end of 3rd grade, the most striking sustained subgroup finding was related to children from high risk households. For this subgroup, children in the 3-year old cohort demonstrated sustained cognitive impacts across all the years from pre-K through 3rd grade. At the end of 3rd grade, the Head Start children from high risk households showed favorable impacts on the ECLS-K Reading Assessment, the WJIII Letter-Word Identification, and the teacher-reported reading/language arts skills. This was in contrast to the impacts for children in lower and moderate risk households, for whom there were no impacts. Those children who started out with more familial stressors than their peers were found to have multiple positive impacts on the direct student assessments over time. Also, among the 3-year-old cohort, children of parents with no reported depressive symptoms experienced sustained benefits of Head Start in the cognitive domain through the end of 3rd grade and in the social-emotional and parenting practices domain through the end of 1st grade.

Among the 4-year-olds, the subgroups that demonstrated sustained benefits are children of parents who reported mild depressive symptoms, severe depressive symptoms, and Black children. Head Start children of parents reporting mild depressive symptoms demonstrated favorable cognitive impacts through the end of 3rd grade. This was in contrast to those with no, moderate, or severe depressive symptoms. However, favorable impacts were reported only at the end of the Head Start year for parents with severe depressive symptoms. In the parenting and social-emotional domains, predominantly favorable parent-reported impacts were sustained for children of parents with severe depressive symptoms. Black children experienced favorable impacts in the social-emotional domain at the end of kindergarten through 3rd grade as reported by teachers, parents, and the child self-report.

Finally, several subgroups experienced solely-or primarily-unfavorable impacts of Head Start that were sustained through 3rd grade. For the 4-year-old cohort, this included White children, who had unfavorable impacts in the social-emotional domain, and for the 3-year-old cohort, children of parents with mild depressive symptoms, who also had unfavorable impacts in the social-emotional domain. Many subgroups in both age cohorts experienced a mixture of favorable and unfavorable impacts, particularly in the social-emotional domain.

---

[15] The Benjamini-Hochberg test of multiple comparisons was also applied to the subgroup analysis, and the results are included in the Chapter 8 tables of this report.

Some of the subgroup impacts from earlier years were not sustained through 3rd grade. For example, the favorable social-emotional impacts for children in the 4-year-old cohort of parents with mild depressive symptoms and the favorable cognitive impacts found at the end of 1st grade for children in the 3-year-old cohort from non-urban settings were not sustained through 3rd grade.

Final Thoughts

Head Start has the ambitious mandate of improving educational and developmental outcomes for children from economically disadvantaged families. Head Start's mandate requires that it meet the needs of the whole child, including the cognitive, social-emotional, and health needs of children, and positively influence the parenting practices of their parents. This study examines the impacts of Head Start on these four domains and whether earlier impacts were sustained into 3rd grade.

The lasting effects of Head Start and early childhood education in general on children's outcomes have been the focus of much study. Considering only outcomes through early elementary school and middle childhood, results for the HSIS cognitive outcomes are in line with other experimental and non-experimental early education studies. Non-experimental Head Start studies showed initial positive impacts of a roughly similar magnitude to those found in the HSIS that dissipated as the children entered early elementary school (Currie & Thomas, 1995; Garces, et al., 2002; Ludwig & Phillips 2008; Deming 2009). Moreover, recent longitudinal data from the experimental evaluation of Early Head Start (Vogel, et al., 2010) showed a similar pattern of early positive impacts that were not sustained into elementary school. Experimental results from the HighScope Preschool Curriculum Comparison study found negligible differences between study groups in cognitive and academic outcomes in the first decade of study (Schweinhart & Weikart, 1997). Similar conclusions about the size and lack of persistence of early impacts were reported in a recent broader meta-analysis of early childhood interventions (Leak et al., 2010). However, as we discuss later, some studies, including those that did not show differences in elementary school, reported finding positive effects later in adulthood. Although the underlying cause of the rapid attenuation of early impacts is an area of frequent speculation, we don't have a good understanding of this observed pattern. All we can say is after the initially realized cognitive benefits for the Head Start children, these gains were quickly made up by children in the non-Head Start group.

We do not yet know if there will be positive outcomes for HSIS participants later in life, however, research suggests that positive outcomes later in life are possible. Despite a growing body of research about relatively rapid dissipation of early cognitive impacts, there is some evidence suggesting that positive effects of Head Start may have an impact on participants' later life such as later school success and early adulthood outcomes (Garces, et al., 2002; Ludwig & Miller, 2007; Deming, 2009). Garces, Thomas, and Currie (2000) conducted a non-experimental study that reported evidence of long-term improvement for Head Start participants on outcomes such as school attainment, earnings and crime reduction, for some race and gender combinations. Ludwig and Miller (2007), using a regression discontinuity design, reported that increases in Head Start funding were associated with a decline in mortality rates for children ages five to nine from causes of death that could be affected by the program, an increase in high school completion, and an increase in the likelihood of attending some college. Both of these

findings were based on Head Start programs that operated in the 1960's through the 1980's. More recently, Gelber and Isen (2011), using the HSIS data, reported that parents of children assigned to Head Start were more involved with them in a variety of activities both during Head Start enrollment and the early elementary years. The authors suggested that increases in parent involvement may mediate long-term impacts on child outcomes. According to a recent paper by Gibbs, Ludwig, & Miller (2011) such delayed or "sleeper" effects may occur because of the Head Start benefits in the area of children's social and emotional development, i.e., improved socialization and emotional strength may have later school-related payoffs.

Research from non-Head Start samples with similar populations also suggests that "sleeper" effects may present years after exposure to early education. Using data from the randomized study, Project Star (1985-89 Tennessee K-3 Class Size Study), Chetty et al. (2010) reported that improvements in kindergarten test scores translate into higher lifetime earnings, more likely college attendance, retirement savings, home ownership, and residence in a better neighborhood. Children from the HighScope group completed more years of school, had less self-reported misconduct at age 15, fewer felony arrests, and fewer property crime arrests than those who received direct instruction. Initially, no early academic differences were found but the long term impacts suggested benefits from quality early childhood education in early adulthood outcomes. Although Project Star and the HighScope Preschool Curriculum Comparison study were not focused on Head Start, like the other Head Start studies, they point to the importance of early education for improving children's long-term outcomes.

In addition to considering the possible long-term impacts, there are a few other things to consider in interpreting the findings of this study. First, this was not a comparison of Head Start to parental care. This study evaluated the Head Start program as it existed in 2002 against a mixture of alternative care settings rather than against a "no services" condition. About 40 percent of the control group did not receive formal preschool education and, for those who did, quality was generally lower than in Head Start. Nevertheless, many of the control group children received services. Further, among those who participated in non-parental care, the control group children were actually in non-parental care for more hours than the Head Start group—on average, children in the control group attended some type of non-parental care about four to five hours more per week in the Head Start year, compared with children who had access to Head Start. Consequently, to achieve measurable impacts, Head Start (as noted above) had to outperform what control group children received.

Additionally, to date the findings do not differentiate impacts for children who received services of differing quality in Head Start. Although the quality was high on average, Head Start programs varied in terms of academic instruction in the key areas measured as part of this study, i.e., early development of language and literacy and mathematics skills. This is not to say that all Head Start programs were not trying their best to improve children's development in these areas, but rather on average the program may not have been potent enough in this particular domain to provide the level of overall learning gains needed to move children into a different, and more rapid, growth trajectory. The pattern for the HSIS data showed initial accelerated gains for the Head Start children, then these gains were quickly made up by the control group children, followed by continued gains at the same pace for both groups. The variation in quality may have contributed to the lack of statistically significant differences in the cognitive domain in the early elementary grades.

A separate report will explore how variation in Head Start quality is related to children's impacts as well as how children's later experiences in the school and community affect their outcomes at 1st and 3rd grade, including whether some later experiences help to sustain impacts through the early elementary grades.

The study also reflects on the impact of Head Start as it existed in 2002, and does not necessarily represent either the impact of Head Start between the time it was initiated and 2002 or the impact of any changes made to Head Start since 2002. As most evaluations, this study is designed to ask a set of questions about a program at a particular point in time. To the extent that the program has changed since the time that study participants were given access, those changes will not be reflected in the study's findings.

Finally, this study leaves many important questions about Head Start unanswered. These questions include, but are certainly not limited to: Is there a benefit to having two years of Head Start rather than one year? What accounts for the subgroup patterns observed in this report? The Head Start Impact Study is an excellent data base for methodological and child development research due to its size, longitudinal data, and multiple variables. Hopefully, researchers will take advantage of the data from this study, which will be made available through a data archive,[16] to further the understanding of the role Head Start plays in the well-being of children and families.

[16] The data is archived at the Child Care & Early Education Research Connections Project. http://www.childcareresearch.org/childcare/welcome

Chapter 1: Study Goals and Purposes

This report is a follow-up to previous reports on the Head Start Impact Study that covered impacts on eligible children and their parents from preschool through the end of 1st grade. This report provides findings on longer term effects through the end of 3rd grade.

The Head Start Program

The Head Start program, created in 1965 as part of the War on Poverty, is intended to help "… preschoolers develop the early reading and math skills they need to be successful in school … by enhancing the social-emotional and cognitive development of children through the provision of educational, health, nutritional, social and other services to enrolled children and families. They engage parents in their children's learning and help them in making progress toward their educational, literacy and employment goals. Significant emphasis is placed on the involvement of parents in the administration of local Head Start programs."[17]

Head Start is administered by local grantees, including public and private non-profit and for-profit agencies, that must adhere to national program guidelines—the Head Start Performance Standards—to ensure that programs provide a wide array of comprehensive services for families and children. Local Head Start programs conduct a needs assessment of each child's and family's needs and strengths at the beginning of the program and use this to tailor and guide services. Local grantees also conduct a community needs assessment that is used to tailor their program options to those that will best fit the needs of the local community. Finally, ongoing assessments of children's development and progress are used to individualize services at the classroom level.

Head Start has grown from its early days of originally offering six-week summer sessions for 4-year-olds, to providing typically nine-month, and sometimes year-long, programs serving children from three to five years of age. In 1995, the Office of Head Start expanded services to low-income pregnant women and families with infants and toddlers through the Early Head Start Program. In general, during the period of this study, to be eligible for Head Start, a

[17] http://www.acf.hhs.gov/programs/ohs/about/index.html.

child had to be living in a family whose income was below the Federal poverty line. Programs were permitted, however, to fill ten percent of their enrollment with children from families that are over this income level. More recently, since the 2007 reauthorization of Head Start, programs may serve up to 35 percent of their enrollment from children whose families' incomes are below 130 percent of the poverty line if slots remain in a program after all interested families at 100 percent poverty have registered. At the time of the study programs were, and still are, required to make at least ten percent of the total number of enrollment opportunities during each enrollment year available to children with disabilities.

The Head Start program offers services to children and families through a variety of program options. The most common of these, and concomitantly the highest proportion of the study sample, is a center-based program option in which children are enrolled in classroom settings and parents participate in at least two home visits annually. Three other options represented in the sample are: (1) a home-based program option in which staff work directly with children and parents primarily in the home on a weekly basis and also in at least twice monthly group socialization activities, (2) a family child care option, in which services to children and families are provided in a family child care setting, and (3) the combination program option that allows for a variety of combinations of center-based class sessions with home visits. Grantees may propose to offer any or all of these options, or may design a different option subject to approval from the Office of Head Start.

As noted above, each program conducts a community needs assessment to determine which options and services best fit the strengths and needs of families in the community. Head Start programs work in partnership with other service providers, adjust schedules to meet the needs of the populations served, vary in length from school-year to full-year, and for those that provide services to children in out-of-home settings, can vary in intensity from part-day to full-day. All of these variations are represented in the sample for this study.

Since 1965, the context in which the program delivers services has changed dramatically. Most notably, greater cultural diversity of the population and increasing prevalence of Dual Language Learners have combined to increase the diversity of children and families, requiring Head Start programs to be responsive to a broader array of ethnic, cultural, and linguistic heritage populations. Increased immigration from Latin America, the Caribbean, Asia,

and the Middle East has created a more diverse population of American children (Cappella & Larner, 1999), and resulted in Head Start serving a wider variety of ethnic and racial minority groups. Since the inception of the program, family structure also has changed with the decline in children living with both parents. For example, births to unmarried women in the U.S. have risen dramatically. In 1980, there were approximately 685,000 births to unmarried women compared to 1,365,966 births to unmarried women in 2002. The proportion of all births to unmarried women was 18.4 percent in 1980 and reached 34 percent in 2002. By 2007, the number rose to 1,714,643 births to unmarried mothers and 39.7 percent of all births to unmarried women (Ventura, 2009). In addition, there has been an increase in the number of children whose parents are divorced (Johnson and O'Brien-Strain, 2000). At the same time, there has been an increase in the number of mothers entering the workforce and seeking child care. Finally, parents have more options for other early childhood care and education. In 2001-2002, there were 693,000 children enrolled in state-funded preschool programs. By 2002-2003, the first Head Start year for the study participants, there were 738,000 children enrolled in state-funded preschool programs (an increase of 45,000 children) (Barnett, Hustedt, Robin, & Schulman, 2004). The enrollment of 4-year-olds in state-funded preschool programs has risen from 14 percent of the national population in 2002 to 28 percent in 2011 (Barnett, Carolan, Fitzgerald, & Squires, 2011).

Similar to the demographic changes, the Head Start program has changed since the first year of the study. In 1965, Head Start enrolled 561,000 children in an eight-week summer program. In contrast to the program in 1965, most children in 2003 (the first year of this study) participated in a part-time, nine-month program during the school year. In addition to moving from a summer program to a school year program, recent changes include (1) the Outcomes Framework that outlines the essential areas of development and learning that are to be used by Head Start programs to establish school readiness goals for their children, monitor children's progress, align curricula, and conduct program planning; (2) the use of the Classroom Assessment Scoring System (CLASS) in program monitoring and professional development; and (3) the Designation Renewal System (DRS) that specifies seven conditions that HHS will consider when determining whether a grantee is delivering a high-quality and comprehensive program and, thus, whether the grantee may be renewed without having to compete for continued funding. These changes should strengthen the ability of Head Start to serve poor and at-risk children in their early years--the program's primary objective since 1965.

At the time this study was initiated, Head Start had grown dramatically and was serving a diverse set of children. According to Head Start data reported by programs to the Federal government, in federal fiscal year 2003, enrollment for the entire Head Start program (including Early Head Start, Migrant and Seasonal and American Indian/Alaska Native tribal programs) was 909,608 children (U.S. Department of Health and Human Services, 2004), with a total of over 22 million preschool-age children, infants, and toddlers having received Head Start services since 1965. Most of the children who received Head Start services in fiscal year 2003 were between three and five years old (92 percent); the remaining (8 percent) were infants and toddlers. At that time, the program included 1,670 Head Start grantees, 47,000 classrooms, 19,200 centers and 206,000 staff, who, with the assistance of 1,372,000 volunteers, worked to provide comprehensive services to meet the early childhood development, educational, health, and family needs of children.

The Congressional Mandate for This Study

The U.S. General Accounting Office (GAO) concluded in the late 1990s that (1) "… the body of research on current Head Start is insufficient to draw conclusions about the impact of the national program" (GAO, 1997) and (2) ". . . 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" (GAO, 1998).

Based on the GAO's recommendations, and on the testimony of research methodologists and early childhood experts, Congress included in the 1998 reauthorization of Head Start a mandate that the U.S. Department of Health and Human Services (HHS) determine the impact of Head Start on the children it serves. The legislation also required the Secretary of HHS to establish an Advisory Committee on Head Start Research and Evaluation. As first established in 1998, the Committee included 30 experts in areas of program evaluation and research, education, child care and early childhood programs, education policy, and economics. In 1999, they set forth a framework for research on the impact of Head Start that would be 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, HHS competitively awarded a contract in October 2000 to Westat of Rockville, MD, and its team of collaborating partners, which currently includes Ronna Cook Associates, Chesapeake Research Associates, LLC, Abt Associates, and the University of Virginia.

Study Objectives and Research Questions

Study Design Requirements

The congressional mandate required a study of the "impact" of Head Start, defining the term in the legislation as the "*. . . difference in an outcome for a participant in a program that would not have occurred without the participation in the program."* Thus, impact means a difference between the outcomes observed for Head Start participants and what would have been observed for these same individuals had they not had the opportunity to participate in Head Start. Although the language in the legislation permitted different study designs, the Advisory Committee concluded that a randomized controlled trial (RCT) would provide the most reliable evidence of causal linkage between Head Start and intended child and family outcomes. As will be discussed below, this was the study's design, in which a sample of 3- and 4-year-old Head Start applicants not previously served by the program[18] were randomly assigned either to a Head Start 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 could receive any other available 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 child and parent outcomes. The advantage of this research design is that if random assignment is properly implemented with a sufficient sample size, program participants on average will not differ in any systematic way from non-participants except through their access to Head Start services.[19] This is true for both measured and unmeasured characteristics of the participants.

---

[18] 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.

[19] 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).

The legislation also suggested that the control group should represent the real world, i.e., it should be *". . . 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."* In other words, the effects of Head Start were to be compared to the range of options that low-income families have for their preschool children and not against a no-services alternative, in which all children would spend all of their time at home with parents or other family members.

Although this type of comparison group does not allow estimation of the impact of Head Start compared with no services, it does allow addressing questions that are relevant for the Head Start program as it currently operates. In using this type of comparison group, the study examines how well the program performs against other alternatives that are currently available for low-income children. A comparison to "stay-at-home" children is not reflective of the real world as it currently exists for most low-income children today.

The legislation also clearly intended that the study be nationally representative, stating that: *"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"* and that the selection of study participants should *". . .make use of random selection from the population of all Head Start programs. . .in selecting programs for inclusion in the research."* Accordingly, the study was designed with a nationally representative, probability-based sample of eligible programs, centers, and children.

Finally, Congress specified that the intent of the study was to examine program impacts over time, i.e., *". . .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). . . ."*

Although not a part of the Congressional mandate, HHS decided to continue the study to examine program impacts at the end of 3rd grade. To date, findings have been reported through 1st grade. This report focuses on the impacts at 3rd grade.

Research Questions

Following the legislative requirements and the Advisory Committee's recommendations, the study was primarily designed to answer questions about the program's overall national impact:

- **The Direct Impact of Head Start on Children***:* What is the impact of Head Start on children's cognitive development preceding the start of school and during the early school years? What is the impact of Head Start on children's social-emotional development preceding the start of school and during the early school years? What is the impact of Head Start on children's health status preceding the start of school and during the early school years?

- **The Potential Indirect Impact of Head Start on Children Through Direct Impacts on Parents***:* What is the impact of Head Start on parents' practices and support of their child's school readiness preceding the start of school and during the early school years as covered in this report, through 3rd grade?

In addition, the study aimed to examine the extent to which impacts vary across different groups of children, parents, and families:

- **Under What Circumstances Does Head Start Achieve the Greatest Impact?** What works for which children? What Head Start services are most related to impact?

  - **Variation by Child Characteristics**: How do the estimated impacts of Head Start vary by child characteristics, such as child's age, primary language, special needs status, and academic ability?

  - **Variation by Parent and Household Characteristics**: How do the estimated impacts of Head Start vary by parent and household characteristics, such as parents' race and ethnicity, depressive symptoms, and level of risk?

  - **Variation by Community Characteristics**: How do the estimated impacts of Head Start vary by the characteristics of the community where the child lived at the time of application to Head Start, such as urbanicity?

  - **Variation by Quality**: How does the estimated impacts of Head Start vary by the quality of the Head Start center? (This question will be addressed in a future report.)

A final research question focuses on the impact of Head Start on the nature and type of children's educational experiences.

- **Impact of Head Start on Children's Educational Experiences**: What is the impact of Head Start on the settings, setting characteristics and services that children experience prior to starting school and during the early school years?

This report addresses the final question in 3rd grade. However, this final question, which requires a more complex set of statistical analyses and assumptions, will be addressed in more detail in a separate report that will examine the relationship between quality in preschool and early school settings and child outcomes.

Contents of Report

This report consists of six chapters and five appendices. Chapter 1 outlines the study goals and purposes, highlighting the Congressional Mandate and research questions. Chapter 2 presents the study design and an overview of the implementation procedures for the 3rd grade follow-up. The degree to which Head Start affected children's school settings, setting characteristics, and the services children received in 3rd grade are discussed in Chapter 3. The main impacts on children's cognitive, social-emotional, and health status are presented in Chapter 4 along with the impacts on parenting practices. Chapter 5 presents the impacts on subgroups and variation in impacts by child characteristics, parent and family characteristics, and community characteristics. Conclusions are discussed in Chapter 6. The appendices include (1) a description of the weighting and analysis procedures that are new for the 3rd grade data, (2) tables presenting detailed information on the baseline characteristics for the 3rd grade analysis sample, (3) the main impact tables for the intent to treat (ITT), (4) the impact tables for the impact on the treated (IOT) and (5) the subgroup impact tables for preschool through 3rd grade.

Interested readers can find the full tables with all details of the analyses conducted for this report on the Administration for Children and Families, Office of Planning, Research, and Evaluation website.[20] This website also includes information from the previous reports on

[20] http://www.acf.hhs.gov/programs/opre/hs/impact_study/index.html.

this study, including the *Head Start Impact Study Final Report* (HHS, 2010),[21] which describes the findings from preschool through 1st grade, and the *Head Start Impact Study Technical Report* (HHS, 2010)[22] which provides details about the sampling methods, psychometric information for the data collection measures, and methods used for collecting and analyzing data.

[21] U.S. Department of Health and Human Services, Administration for Children and Families (January, 2010). *Head Start Impact Study: Final Report.* Washington, DC. http://www.acf.hhs.gov/programs/opre/hs/impact_study/reports/impact_study/hs_impact_study_final.pdf.

[22] U.S. Department of Health and Human Services, Administration for Children and Families (January 2010). *Head Start Impact Study: Technical Report. Washington, DC.* http://www.acf.hhs.gov/programs/opre/hs/impact_study/reports/impact_study/hs_impact_study_tech_rpt.pdf.

Chapter 2: Study Design and Implementation

This chapter describes the design and implementation of the Head Start Impact Study, including procedures used to select the study sample and randomly assign eligible children, characteristics of the study sample, data collection procedures, and analysis methods used to derive the impact findings found in subsequent chapters. More detailed information can be found in the series of reports released in 2010, covering results through the end of 1st grade.[23]

Study Design Overview

As discussed in Chapter 1, the primary purpose of the Head Start Impact Study was to determine whether Head Start, nationally, has short- and long-term impacts on participating children and their parents and the extent to which there is variation in program effects for different types of children and families. To answer these questions required the design and implementation of a unique study:

- **Randomized Controlled Trial (RCT) Design:** The Congressional mandate for the Head Start Impact Study had a clearly stated goal of producing causal findings, i.e., the purpose was to determine if access to Head Start caused better developmental and parenting outcomes for participating children and families. The basic study design involved the random assignment of a sample of 3- and 4-year-old Head Start applicants not previously served by the program,[24] either to the Head Start group or to a non-Head Start control group. This procedure ensured comparability between the two groups at program entry, so later differences can be causally attributed to having access to Head Start. The Head Start group was allowed to enroll in Head Start, while the control group was not granted access to Head Start, but may have received similar services through other available programs chosen by their parents during the first year of the study. To be randomly assigned, the child's eligibility for admission to the program had to have been determined by the local Head Start agency. Thus all children in the study were determined to be eligible for Head Start, regardless of whether they were assigned to the Head Start or control group.

[23] U.S. Department of Health and Human Services, Administration for Children and Families (January, 2010). *Head Start Impact Study. Final Report.* Washington, DC.

U.S. Department of Health and Human Services, Administration for Children and Families (January, 2010). *Head Start Impact Study. Technical Report.* Washington, DC.

[24] 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.

- **Nationally-Representative Sample:** Most random assignment studies are conducted in small demonstration programs or in a small number of operating sites, usually those that volunteer to be included in the research. In contrast, the study sample of Head Start programs, and newly entering 3- and 4-year old children, was selected to be nationally representative of the complete program (with a few exclusions). The sample is, therefore, generalizable nationally to the vast majority of programs and is not a typical "convenience" sample of cooperative programs.

- **Collection of Data on Multiple Outcomes:** The study collected data from children and parents that covered four key domains: cognitive outcomes related to language, literacy and mathematics, social emotional development, health, and parenting practices that support children's school readiness. The study examines impacts in these areas for the 3- and 4-year-old cohorts separately by examining the difference in outcomes between children assigned to the Head Start group and those assigned to the control group within each of the two age cohorts.

- **Longitudinal Study Sample:** The study collected data on the complete sample of randomly assigned study children, and their respective parents, from the time of entry into Head Start in Fall 2002 through the end of 3rd grade (Spring 2007 for the 4-year-old cohort and Spring 2008 for the 3-year-old cohort). To allow for the possibility of a future data collection, the sample was tracked through the end of 2011 and HHS has awarded a contract to Westat to continue tracking the study sample through the end of 2016.

Details on each aspect of the study are provided in this chapter. The first section provides information on sample selection, random assignment, data collection and data sources, and response rates. The next section presents the description of the cognitive, social-emotional, health, parenting practices, and school characteristic measures used in the 3rd grade analysis. The last section of this chapter discusses analysis methods and the presentation of study findings.

Sample Selection

First-time applicants to Head Start in fall 2002 were randomly selected from a nationally representative sample of Head Start programs,[25] making the results generalizable to the national Head Start program as it existed in 2002 with a few exceptions. The study sample did, however, exclude the Migrant and Seasonal Head Start programs, Early Head Start (which serves pregnant women and children from birth to age three), and the American Indian/Alaska Native tribal programs. The study also excluded programs in communities that did not have more

[25] Certain exclusions were made from the universe of all Head Start grantees in the country, for reasons described in the *Head Start Impact Study Technical Report*.

children eligible for Head Start than could be served with the existing number of funded slots. This constraint ensured that the study's need for a control group did not require slots to go unfilled.[26] The study used a multi-stage sampling process to select a representative group of Head Start programs. The process, depicted in Exhibit 2.1, is described below:

1. **Identify Grantee/Delegate Agencies.** 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) 1998-99, after excluding (1) grantee/delegate agencies serving only special populations (migrant/seasonal and tribal Head Start programs and sites serving only Early Head Start children), (2) grantees involved in the FACES 2000 study, and (3) as recommended in the Advisory Committee report (1999), grantees/delegate agencies that were "*extremely new to the program*."[27]

2. **Create, Stratify, and Select Geographic Clusters.** This pool of 1,715 Head Start programs was subsequently organized into 161 "geographic clusters" (to increase our ability to closely monitor random assignment and obtain high-quality data). The clusters were then grouped into 25 strata to ensure variation in 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. This resulted in 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).

3. **Determine 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." In order to address Head Start grantee concerns about denying children access to Head Start and to better ensure grantee participation, 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 reduces the ability to generalize the results to some extent, as discussed later in the chapter. Eligibility was determined from information verified through telephone calls to all initially sampled 261 grantee/delegate agencies, augmented with information provided by Federal Regional Office staff and with data obtained from secondary sources such as local

[26] Taking into account all of these opportunities for Head Start grantees/delegate agencies and centers to be excluded, the estimated weighted national coverage rate for spring 2003 data was 84.5 percent, meaning that the study sample was representative of 84.5 percent of the total universe of all newly entering 3- and 4-year-olds across the country. (The small number of grantees/delegate agencies and centers that was found to be closed or merged into another program or center was considered as ineligible, as they were no longer part of the universe from which the sample was drawn.)

[27] Defined as in operation for fewer than two years.

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, ten other grantee/delegate agencies had been closed or merged, further reducing the pool of eligible programs to 223 grantee/delegate agencies.

4. **Stratify and Select Grantee/Delegate Agencies.** Under a PPS (Probability Proportional to Size) sample design, the largest programs have the highest probability of being selected. 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." The single grantee/ delegate agencies, and the formed groups, were then stratified along several dimensions to ensure that programs selected represented the following conditions: 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 agencies or 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 agencies or groups comprising 90 individual grantee/delegate agencies across 23 states.

5. **Recruit Grantee/Delegate Agencies.** Senior project staff visited all 90 selected grantee/delegate agencies during summer 2001 to explain the study, verify information needed for study implementation, and 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 Centers[28] Consortium—leaving 87 grantee/delegate agencies in 76 grantee/delegate agency groups (i.e., the overall number of grantee/delegate agency groups was not reduced).

6. **Develop List of 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.

7. **Determine Eligible Centers and Create Center Groups.** 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

[28] The Head Start Bureau (HSB) 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 promising approaches to promoting the school readiness of Head Start children.

grantee/delegate agency groups (a reduction of 11.8 percent and the loss of three grantee/delegate agencies, but no grantee groups). Next, small centers were combined with nearby centers to create "center groups."

8. **Stratify and Select a Sample of Study Centers.** 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). Three center groups were 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 in 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[29] to yield a final sample of 383 Head Start centers.

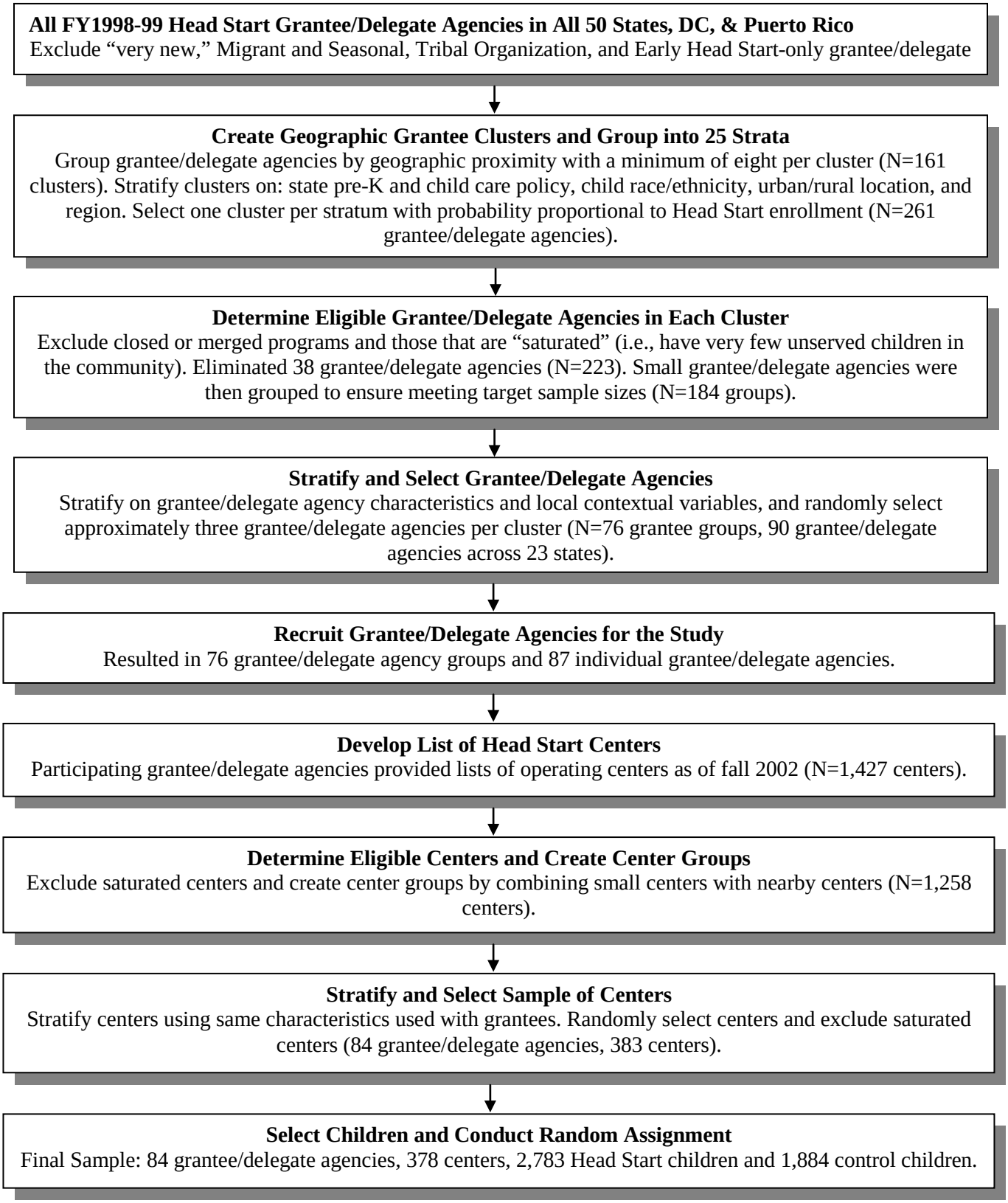
9. **Select Children and Conduct Random Assignment.** The sample of Head Start grantee/delegate agencies and centers, when properly weighted, was designed to yield a sample of children that represented the national population of newly entering children and their families (with the exclusions noted above) for the 2002-03 program year. The sample of children included 2,783 Head Start children and 1,884 control children. Details on random assignment are described below.

A complete discussion of sampling and weighting is provided in the *Head Start Impact Study Technical Report.*[30] Appendix A presents updated weighting and analysis procedures for the 3rd grade follow-up data.

[29] 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.

[30] See Footnote 23 in this chapter.

Exhibit 2.1. Sample Selection Process for the Head Start Impact Study



Random Assignment

At each of the selected Head Start centers, program staff provided study information 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. 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 might be admitted if eligible.

Study staff obtained data on all applications for the 2002-03 program year (to ensure that all applicants were considered for random assignment). Returning children, and a small number of grantee-requested "high-risk" exclusions,[31] were eliminated from consideration for the study. 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 assess these children's outcomes for the study (e.g., blindness). Each grantee was limited to one exclusion per center. There were 276 exclusions granted prior to the random assignment of the children.

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. Study staff 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 control group from who would normally be next in line for admission. Study children were randomly selected from the entire list.

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

[31] This decision was made because there were concerns about assigning very high-risk children to the control group, especially in situations where Head Start may provide their only option for early childhood services.

be assigned to the control group. For an average center group, the 11 control group children represented about nine percent of total enrollment. In some cases, where fewer children than expected were actually available, a smaller sample of children was selected for the study.

The study was designed to separately examine two cohorts of children, newly entering 3- and 4-year-olds. This design reflects the hypothesis that different program impacts may be associated with different age of entry into Head Start. Differential impacts are of particular interest in light of 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. Consequently, the study included two separate samples: a newly entering 3-year-old group[32] (to be studied through two years of Head Start participation, kindergarten, 1st grade, and 3rd grade) and a newly entering 4-year-old group (to be studied through one year of Head Start participation, kindergarten, 1st grade, and 3rd grade). The 3-year-old group was slightly larger than the 4-year-old group to protect against the likelihood of higher study attrition resulting from an additional year of longitudinal data collection for the younger children.[33] The 3-year-old cohort was randomized for only one year of access to Head Start. During their second potential Head Start year, by study design, the 3-year-old control group cohort was given access to Head Start. The interpretation of impacts on the 3-year-old cohort, given a single year of randomization is discussed later in this chapter.

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.[34] In total, 4,667 newly entering children were randomly assigned and included in the Head Start Impact Study (see Exhibit 2.2).

---

[32] Newly-entering indicates that the child has not been previously enrolled in Early Head Start or Head Start.

[33] This roughly equal sampling of 3- and 4-year-old applicants was done to obtain reliable estimates of program impacts for each age cohort, 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 total of all 4-year-old participants includes 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.

[34] The five centers were excluded due to center closures and mergers.

Exhibit 2.2. Number of Children Randomly Assigned to Head Start and Control Groups, by Age Cohort

Age Cohort	Head Start Group	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 control group. This imbalance reduces the precision of the impact estimates by less than two 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) the loss of sites due to lack of excess demand was decreased, and (3) the cost of data collection was decreased because Head Start group members require less effort to track and interview over time than children in the control group.

The Success of Random Assignment

A comparison of demographic characteristics of the randomly assigned children and their parents indicated that there were few statistically significant differences[35] between the Head Start and control groups. 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.

Random assignment rarely results in perfect adherence to the assigned program status. In the current study, one would expect some children assigned to the Head Start group not to participate 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"). During program recruitment, Head Start grantees and centers described no-shows as a common occurrence in ordinary program operations, with rates among enrolled children often in the double digits.

[35] Among 16 variables (e.g., child gender, child ethnicity, etc.) collected at baseline, there were differences in very few. For the 3-year-old cohort, there was a significant difference on the parent/caregiver age (Head Start parents/primary caregivers were 0.9 years older, on average, than control group parents/primary caregivers) and a grandparent was more likely to live in the household for the Head Start group than the control group. For the 4-year-old cohort, Head Start group mothers were more likely to report education beyond high school than control group mothers and Head Start group households were less likely to receive TANF than control group households. See Exhibits 2.9-A and 2.9-B in the *Head Start Impact Study Final Report* for a listing of all the variables.

Consequently, it is not surprising that some families who were randomly assigned to the Head Start group subsequently opted for a different care setting for their child.[36]

Similarly, although every effort was made to maintain the integrity of the control group, perfect conditions could not be implemented. In some instances, local staff intentionally enrolled control group children into Head Start. More commonly, parents simply applied to another nearby Head Start program, especially in densely populated areas with Head Start programs operating in proximity. Due to confidentiality restrictions, information on study participants was not shared with programs not involved in the study, so control group families were not prevented from being served by other Head Start programs.

For analysis purposes (as explained below), it is only the degree of compliance with the random assignment design *in the first year of the study* that matters, since this was the one year in which the study sought to have all Head Start group children—and none of the control group children—participate in Head Start. Exhibit 2.3 provides information on the incidence of Head Start group no-shows and control group crossovers by age group in that year. In the exhibit, children in the Head Start group were considered no-shows if it was determined that they did not participate in Head Start at any time during the 2002-03 program year. Children in the control group were deemed crossovers if they participated in Head Start at any time during the 2002-03 program year. This determination was based on information from parent surveys, checking Head Start enrollment in fall 2002, and the care setting identified at the time of the child's fall and spring assessments. No-shows accounted for 15 and 20 percent of the full randomly assigned Head Start samples for children in the 3- and 4-year-old cohorts, respectively; crossovers accounted for 17 and 14 percent of the randomly assigned control groups.

[36] See the *Head Start Impact Study. Final Report* (January 2010) for additional information.

Exhibit 2.3. The Incidence of No-Show and Crossover Behavior for the Sample as Randomly Assigned, by Age Cohort (Weighted Data)

Sample Group	Some Year 1 Head Start Participation	No Year 1 Head Start Participation	Total
All Randomly Assigned (N=4,667):			
3-Year-Old Cohort			
Head Start Group	85.1%	14.9%	100%
Control Group	17.3%	82.7%	100%
4-Year-Old Cohort			
Head Start Group	79.8%	20.2%	100%
Control Group	13.9%	86.1%	100%

Data Collection and Data Sources

Data collection began in fall of 2002[37] and continued through the spring of 2008, following children from entry into Head Start through the end of 3rd grade. Data collection included the following components.

- **Direct Child Assessments.** The child assessment battery administered annually focused on language and literacy, including children's vocabulary knowledge, reading and writing skills and achievement, oral comprehension and phonological awareness, and math skills and achievement. The 45- to 60-minute child assessment battery was typically administered one-on-one by specially trained assessors in the child's home during the elementary school years. The 3rd grade child assessment included direct measures of the child's ability in reading and mathematics.

- **Child Survey.** The child survey is a self-assessment of the child's academic and social skills and was administered at the same time as the child assessment at 3rd grade. It was administered using a Touch Screen on a laptop computer for responses and earphones to listen to the questions.

- **Parent Interviews.**[38] In-person interviews were typically conducted in the home of each study child with a parent or primary caregiver living with, and responsible for raising, the child at the fall 2002 baseline point and at each of the subsequent spring data collections through the child's 3rd grade year. It was possible that the parent or primary caregiver could change over time, but this occurred for a very small percentage of the children. Parent interviews were available in English and Spanish

[37] Fall 2002 data collection was completed between the end of September and mid-November for the majority of children and parents (although a small number did extend into December). The discussion of analysis procedures in this chapter and in the *Head Start Impact Study Technical Report* (2010) detail how this late baseline data collection is handled in the analysis of program impacts.

[38] In addition, in the winter of 2003, and in the fall of each subsequent year, a 10-minute telephone interview was conducted with the parent/primary caregivers to obtain up-to-date contact information and information regarding the child's current preschool, child care, or school placement to determine the appropriate setting for the spring data collection waves. If parents could not be reached by telephone, in-person interviews were conducted to collect this information.

versions, and bilingual English/Spanish speakers were hired for areas with Spanish-speaking families. For other languages, interviewers/assessors fluent in these languages were hired or local resources were asked to identify interpreters to aid in completing the parent interviews.

Information collected from parents during the interviews included: (1) parents' report of a variety of child-specific information, including the child's demographic characteristics, behavior, developmental accomplishments, disabilities, and child care arrangements; (2) parental characteristics such as education, employment, and reported depressive symptoms; (3) household characteristics, such as household risk, household members and income; (4) parent-child activities and interactions such as going to the library with the child; (5) parenting practices such as safety practices and parenting styles; (6) the child's experiences during preschool and early elementary school years, including parent communication and involvement with school; and (7) community characteristics such as crime in the neighborhood.

- **Teacher Surveys.** Teachers were asked to complete a survey in spring of each data collection period from preschool through 3rd grade. During the preschool years, there was no teacher survey for children who remained at home with their parents and did not participate in any out of home care. If, during the preschool years, a child was in a care setting other than Head Start or other center-based programs (e.g., day care home), a survey comparable to the teacher survey was completed by the care provider. Similar to the earlier teacher surveys, the 3rd grade teacher survey included questions about teacher characteristics (e.g., teacher training and education), the classroom environment (e.g., number of children, race/ethnicity of children in the class, number of children eligible for free- or reduced price lunch, classroom attendance and behavior), and the type and frequency of language and math activities used in the classroom. Additionally, the 3rd grade teacher survey included questions on professional development, participation in school decisions, teacher collaboration, teacher satisfaction, and use of computers.

- **Teacher Child Reports (TCRs).** Teacher ratings of children's accomplishments and behavior are an important source of information about children's learning and behavior because teachers see children over extended periods of time in different settings, providing for appraisals of children's skills and competence in those settings. Moreover, these reports can be important in and of themselves because they reflect the way these individuals interact with the children. The teacher ratings at all data collection points provided measures of the teacher child relationship, the child's behavior, and the child's classroom performance. In kindergarten, 1st grade, and 3rd grade, teachers were asked general questions about the child (e.g., promotion to the next grade), rating of their overall skills in language and literacy, mathematics, and science and social studies, and rating of their school accomplishments. The 3rd grade teacher survey included more detailed questions on reading/language arts and mathematics classroom characteristics such as the amount spent each day on the subject, the number of nights per week the child has homework in the subject, and the types of classroom activities (e.g., discuss new or difficult vocabulary, explain how a math problem is solved, etc.).

- **Principal Surveys**. The principal survey was added for the 3rd grade data collection. Principals were another source of data for school demographic characteristics and quality indicators for the school, teachers, and classrooms. The principal survey included questions on school characteristics, faculty and staff, instruction, technology, decision-making, educational goals, and community relations. The principal also was asked to complete a rating of each study child's reading and/or math teacher.

Response Rates

Exhibits 2.4 (4-year-old cohort) and 2.5 (3-year-old cohort) present response rates[39] for all study instruments administered over the entire data collection period from fall 2002 through spring 2008. Some instruments were not administered in all data collection periods (these instances are indicated as NA). Parent and child assessment response rates represent the number of interviews/assessments completed, i.e., the percentage of the randomly assigned sampled population that completed at least 90 percent of the items for the interview or assessment. For each year, the response rate is calculated on the entire randomly assigned sampled population, not just on those with completed responses the previous year.

Exhibit 2.4. Treatment and Control Response Rates for All Study Years and for All Study Instruments for the 4-Year-Old Cohort

Instruments	Fall 2002 (Baseline)		Spring 2003 (End of Head Start Year)		Spring 2004 (Kinder-garten)		Spring 2005 (1st Grade)		Spring 2007 (3rd Grade)	
	T	C	T	C	T	C	T	C	T	C
Child Assessment	86%	77%	87%	77%	81%	74%	79%	73%	72%	71%
Parent Interview	90%	84%	85%	79%	82%	75%	82%	75%	77%	75%
Teacher Survey	NA	NA	90%	70%	64%	68%	78%	81%	63%	60%
Teacher Child Reports	NA	NA	90%	70%	64%	68%	78%	81%	63%	60%
Principal Survey	NA	NA	NA	NA	NA	NA	NA	NA	66%	63%
Center Director Interviews	NA	NA	91%	73%	NA	NA	NA	NA	NA	NA
Classroom Observations	NA	NA	92%	68%	NA	NA	NA	NA	NA	NA

[39] For additional information on response rates, see the *Head Start Impact Study Final Report* (2010) and the *Head Start Impact Study Technical Report* (2010).

Exhibit 2.5. Treatment and Control Response Rates for All Study Years and for All Study Instruments for the 3-Year-Old Cohort

Instruments	Fall 2002 (Baseline)		Spring 2003 (Head Start Year)		Spring 2004 (Age 4 Year)		Spring 2005 (Kinder-garten)		Spring 2006 (1st Grade)		Spring 2008 (3rd Grade)	
	T	C	T	C	T	C	T	C	T	C	T	C
Child Assessment	87%	76%	89%	80%	87%	79%	82%	77%	81%	74%	79%	71%
Parent Interview	93%	84%	88%	81%	86%	79%	85%	79%	85%	76%	83%	75%
Teacher/Care Provider Survey	NA	NA	88%	64%	87%	79%	82%	84%	86%	88%	63%	60%
Teacher/Care Provider's Child Reports	NA	NA	88%	64%	87%	79%	82%	84%	86%	88%	63%	60%
Principal Survey	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	66%	58%
Center Director Interviews	NA	NA	86%	81%	78%	73%	NA	NA	NA	NA	NA	NA
Classroom Observations	NA	NA	91%	66%	87%	84%	NA	NA	NA	NA	NA	NA

Measures

The 3rd grade measures[40] described in this section fall into three categories: (1) child and family demographics collected at baseline that were used as covariates in the impact analyses, (2) child and family outcome measures, and (3) characteristics of the 3rd grade schools and classrooms attended by the study children. Each area is described below with details provided on individual measures used in these analyses. Citations for published measures are included in the Reference section of this report

Child and Family Demographics

The following measures were created from data collected at baseline and used as covariates, and to create subgroups, for the impact analysis: child's race/ethnicity, child's gender, child's baseline academic skill level, whether biological father lives with child, whether grandparent lives in household, number of adults over age 18 in the household, number of children under age 6 in household, home language spoken, whether the family moved in the last 12 months, family income level, whether the family had economic difficulty in the past three months (paying rent, paying electric and heating bills, buying food for the family, buying clothes for the children), father's and mother's employment status, whether biological mother is a recent immigrant, mother's age, whether mother was a teenager at time of birth of study child, mother's marital status, mother's highest level of education, and mother's reported depressive symptoms. Characteristics that could change also are collected in subsequent study years (e.g., household enumeration, employment status, education level, depressive symptoms, etc.).

Child and Family Outcome Measures

Outcome measures were developed in four domains—child cognitive development, child social-emotional development, health, and parenting practices. The selection of these domains was guided by several factors. First, it was important to measure the school readiness skills that are the focus of the Head Start program. The Head Start performance measures and conceptual framework (U.S. Department of Health and Human Services, 2001) indicate that children enrolled in Head Start should demonstrate improved emergent literacy, numeracy, and

[40] All Cronbach alphas reported for the measures are based on the study sample.

language skills. The framework also stresses that children should demonstrate positive attitudes toward learning and improved social and emotional well-being, as well as improved physical health and development.

Second, domains were selected to reflect the program's whole child model, i.e., school readiness is considered to be multi-faceted and comprising five dimensions of early learning: (1) physical well-being and motor development, (2) social and emotional development, (3) approaches toward learning, (4) language usage, and (5) cognition and general knowledge (Kagan, Moore, & Bredekamp, 1995). The whole-child model also was recommended by the Goal One Technical Planning Group of the National Education Goals Panel (Goal One Technical Planning Group, 1991, 1993).

Third, in 2002, the National Institute of Child Health and Human Development (NICHD), the Administration for Children and Families (ACF), and the Office of the Assistant Secretary for Planning and Evaluation (ASPE) within the U.S. Department of Health and Human Services (HHS) convened a panel of experts to discuss the state of measurement and assessment on early childhood education and school readiness in the cognitive and social emotional domains. Language, early literacy, and mathematics were the primary cognitive domains identified by the experts as important to early childhood development. The experts identified social-emotional competency and regulation of attention, behavior, and emotion as critical measures in the social-emotional domain.

Based on these factors and advice from the experts consulting with the Head Start Impact Study team and the Advisory Committee on Head Start Research and Evaluation, measures were selected to assess the cognitive, social-emotional, and health outcomes of children. Considering the major emphasis Head Start places on parent education and involvement, and its importance for promoting children's development, a fourth domain, parenting practices, was also included. Exhibits 2.6 and 2.7 provide the measures used in pre-K through 3rd grade and the year in which they were administered. The 3rd grade measures are summarized in more detail within this chapter, organized by the four domains. A summary of the

measures used in pre-K through 1st grade is provided in the *Head Start Impact Study Final Report.*[41]

Exhibit 2.6. Cognitive Domain Measures From Direct Assessment and Teacher Report and Year That Measure Was Administered

Construct & Test	What Is Measured	Year Measured			
		Pre-K	K	1st Grade	3rd Grade
Cognitive Domain					
Vocabulary					
Peabody Picture Vocabulary Test III (PPVT adapted)	Vocabulary knowledge and receptive language	X	X	X	X
Color Names	Color identification	X			
Test de Vocabulario en Imágenes Peabody (TVIP adapted)	Vocabulary knowledge and receptive language	X	X	X	
Oral Comprehension					
Woodcock-Johnson III Oral Comprehension	Oral comprehension using syntactic and semantic clues	X	X	X	
Phonetic Awareness					
Preschool Comprehensive Test of Phonological and Print Processing: Elision (CTOPPP)	Phonetics of words, syllables, and phonemes	X	X		
Woodcock-Johnson III Word Attack	Phonetic and structural skills		X	X	
Pre-Writing					
McCarthy Draw-A-Design	Perceptual motor skills	X			
Pre-Reading/Reading					
Letter Naming	Ability to recognize letters of the alphabet	X	X		
Woodcock-Johnson III Letter-Word Identification	Letter and word identification skills	X	X	X	X
Batería R Woodcock-Muñoz Identificación de letras y palabras	Letter and word identification skills	X	X	X	X
Woodcock-Johnson III Passage Comprehension	Word recognition and reading comprehension using syntactic and semantic clues			X	
Woodcock-Johnson III Spelling	Early writing and spelling	X	X	X	

[41] See Footnote 23.

Exhibit 2.6. Cognitive Domain Measures From Direct Assessment and Teacher Report and Year That Measure Was Administered (continued)

Construct & Test	What Is Measured	Year Measured			
		Pre-K	K	1st Grade	3rd Grade
Woodcock-Johnson III Writing Samples	Writing			X	
ECLS-K Reading Assessment	Reading comprehension				X
Math					
Woodcock-Johnson III Applied Problems	Analyze and solve math problems	X	X	X	X
Woodcock-Johnson III Quantitative Concepts – Concepts and Number Series	Knowledge of math concepts, symbols and vocabulary, counting, identifying numbers and shapes, and identifying number patterns		X	X	
Woodcock-Johnson III Calculation	Mathematical computations			X	X
Counting Bears	One-to-one correspondence	X			
School Performance					
Grade Promotion	Parent report of grade promotion		X	X	X
Academic Ratings	Teacher rating of academic ability and skills		X	X	X

Exhibit 2.7. Social-Emotional, Health, and Parenting Practices Domain Measures From Parent and Teacher Report and Year That Measure Was Administered

Construct & Test	What Is Measured	Year Measured			
		Pre-K	K	1st Grade	3rd Grade
Social-Emotional Domain					
Problem Behaviors					
Adapted Child Behavior Checklist	Parent rates child on total problem behavior, hyperactive behavior, aggressive behavior, and withdrawn behavior	X	X	X	X
Strengths and Difficulties Questionnaire	Teacher rates child on emotional symptoms, conduct problems, hyperactivity, peer problems, and pro-social skills				X
Social Skills and Social Competencies					
Developing Skills Checklist	Parent rates child on a number of social skills	X	X	X	
Social Skills and Positive Approaches to Learning	Parent rates child on social skills and learning behaviors	X	X	X	X
Social Competencies	Teacher rates child on social behavior and self-regulation				X
Child Self-Assessment on Academic and Social Skills	Child rates self on internalizing, externalizing, peer relations, and school measures				X
Approaches to Learning					
Adjustment Scales for Pre-School Intervention (ASPI)	Teachers rate child on emotional and behavioral adjustment in the classroom	X	X	X	
Parent-Child Relationships and Teacher-Child Relationships					
Parent-Child Relationship Scale	Parent rates child on closeness, conflict, and positive relationship measures	X	X	X	
Teacher-Child Relationship Scale	Teacher rates child on closeness, conflict, and positive relationship measures	X	X	X	X
Health Domain					
Receipt of Health Care Services	Parent report of child's use of and access to dental care and health insurance	X	X	X	X

Exhibit 2.7. Social-Emotional, Health, and Parenting Practices Domain Measures From Parent and Teacher Report and Year That Measure Was Administered (continued)

Construct & Test	What Is Measured	Year Measured			
		Pre-K	K	1st Grade	3rd Grade
Health Status	Parent report of child's overall health status, child needs ongoing care, child has had care for injury	X	X	X	X
Parenting Practices Domain					
Educational Activities with Child	Parent reports on frequency of reading to child at home, cultural enrichment activities done with child, summer activities	X	X	X	
Discipline Practices	Parent reports on use of physical discipline and use of timeout	X	X	X	X
Safety Practices	Parent reports on the use of 10 safety precautions from seat belts to supervising during bath time	X	X		
Parenting Styles	Parent reports on the degree to which warmth and control exist in parent-child relationship	X	X	X	X
School as Supportive Environment	Parent rating of school on safety, meeting child's needs, and confidence in the school				X
Effect of Parenting on Parent's Life	Parent rating of their life as a parent				X
School Services	Parent rating of communication between the school and themselves				X
Doing Things Together	Parent reports on family activities				X
Time Spent with Child	Parent reports on time with child and time spent in activities with child				X
School Contact and Communications and Parent Participation	Teacher reports on parent and teacher communication and parent participation in school activities		X	X	X

Cognitive Domain[42]

The cognitive test battery consisted of both standardized tests developed by recognized test publishing companies and non-standardized tests developed for use in the Head Start Family and Child Experiences (FACES) or Early Childhood Longitudinal Study – Kindergarten (ECLS-K) projects. As the children aged from Head Start entry through 3rd grade, new tests were added to the child assessment battery; existing tests were extended to include more difficult items; and, in some cases, preschool-level tests were dropped as the children entered elementary school. Each of the 3rd grade cognitive measures, including teacher-reported measures related to reading/language arts and math skills and achievement and parent report of promotion, is described briefly below.

Language and Literacy: Vocabulary

- **Peabody Picture Vocabulary Test (PPVT), Third Edition.** The PPVT measures receptive vocabulary, i.e., listening comprehension for the spoken word in standard English. The child is instructed to look at four pictures and point to the one best representing the meaning of the stimulus word presented orally by the assessor. An adaptive shorter version of the PPVT was used at all data collection points. The adaptive version was first used in the 1997 FACES project. The *Head Start Impact Study Technical Report* provides further details on this adaptation. The Cronbach Alpha for the 4-year-old cohort is 0.81 and for the 3-year-old cohort is 0.80.

Language and Literacy: Reading

- **Woodcock-Johnson III Tests of Achievement: Letter-Word Identification.** The Letter-Word Identification test measures letter and word identification skills. The initial items involve symbolic learning or the ability to match a rebus (pictographic representation of a word) with an actual picture of the object. The remaining items measure a child's reading identification skills in identifying isolated letters and words as they appear in the test easel. The Cronbach Alpha for the 4-year-old cohort is 0.78 and 0.80 for the 3-year-old cohort. The Batería-R Woodcock-Muñoz Pruebas de aprovechamiento-Revisada Identificación de letras y palabras (Cronbach Alpha is 0.83 for both age cohorts) was used for the Spanish and bilingual test administration.

- **Early Childhood Longitudinal Study-Kindergarten (ECLS-K) 3rd Grade Reading Assessment.** The 3rd grade reading assessment emphasized reading comprehension, with the majority of questions based on one of several reading passages. Additional questions tapped basic skills, including decoding and vocabulary. Children began the reading assessment with a routing test of 15 items, five of which were based on a short reading selection. The score on the routing test

[42] For all cognitive outcomes at the 3rd grade level, a higher score indicates better functioning.

was used to select one of three second-stage forms, of varying difficulty, each consisting of four (low form) or five (middle and high forms) reading passages with associated questions, plus five or six individual decoding vocabulary items. The Cronbach Alpha is 0.80 for both age cohorts.

Math

- **Woodcock-Johnson III Tests of Achievement: Applied Problems.** This test measures the child's ability to analyze and solve practical math problems. To solve the problems that are read by the assessor to the child, the child must recognize the procedure to be followed and then count and/or perform simple calculations. The Cronbach Alpha for the 4-year-old cohort is 0.71 and 0.73 for the 3-year-old cohort.

- **Woodcock-Johnson III Tests of Achievement: Calculation.** This test measures the ability to perform mathematical computations. The initial items require the child to write single numbers. The items progress in difficulty from basic operations to geometric, trigonometric, logarithmic, and calculus operations. The calculations involve whole numbers, percents, fractions, decimals, and negative numbers. The Cronbach Alpha is 0.83 for both age cohorts.

School Performance Measures

The tests included in the direct child assessment battery are described above. Other measures of children's cognitive skills included the following:

- **Teacher Report of Academic Ability.** Each child was rated on two academic skills (language and literacy and mathematical skills) by his/her teacher. The child was rated as compared to other children at the same grade level using a five-point scale ranging from one (far below average) to five (far above average). For the analysis, the scores were collapsed to zero (far below average and below average) and one (average, above average, and far above average).

- **Teacher Report of Reading/Language Arts and Math Skills.** Each child was rated by his/her teacher on a series of items that described the child's skills, knowledge, and behaviors focusing on language and literacy and mathematics. The child was rated using a five-point scale that reflected the degree to which the child had acquired the demonstrated skills, knowledge, and behaviors, ranging from one (not yet) to five (proficient). Sample reading accomplishments include "Reads fluently" and "Reads third grade books (fiction) independently with comprehension." Sample math accomplishments include "Uses a variety of strategies to solve math problems" and "Shows understanding of place value with whole numbers." Both the reading/ language arts and math scales were scored using Item Response Theory (IRT).[43] The Cronbach Alpha for the 10-item reading/language arts scale is 0.95 for both age cohorts and for the 10-item math scale is 0.94 for both age cohorts.

[43] These scores are reported in the Item Response Theory (IRT) metric. Since there is not a normalizing sample to link the scores to, the scores are reported in a standardized format with a mean of 0 and a standard deviation of 1.

- **Parent Report of Promotion.** Parents were asked the grade level of their child. This information was compared to the previous year to determine whether or not the child was promoted.

Social-Emotional Domain[44]

Measures for this domain included parents' report of child behavior, teachers' report of classroom behavior and teacher and child relationships, and children's report of their academic and social skills.

- **Social Skills and Positive Approaches to Learning.** Parents were asked to rate their child's social skills and positive approaches to learning. The measure assessed social skills focused on cooperative and empathic behavior and children's approaches to learning such as curiosity, imagination, openness to new tasks and challenges, and having a positive attitude about gaining new knowledge and skills. Examples of the questions asked included: "Makes friends easily," "Comforts or helps others," "Accepts friends' ideas in sharing and playing," "Enjoys learning," "Likes to try new things," and "Shows imagination in work and play." The scale included seven items, with each item scored from zero (not true) to two (very true), and the scale scores can range from zero to 14. The scale was based on an instrument used in FACES and is based on a modified Achenbach Classroom Behavior Checklist (CBCL) (Achenbach, Edelbrock, and Howell, 1987). The Cronbach Alpha is 0.64 for the 4-year-old cohort and 0.61 for the 3-year-old cohort.

- **Problem Behavior of Children.** 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 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. Four scales derived from the instrument are described below. These scales were also used in FACES 2000.

  - The **Total Behavior Problem** scale, derived from parent ratings, included 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 Cronbach Alpha is 0.80 for the 4-year-old cohort and 0.82 for the 3-year-old cohort.

  - The **Aggressive Behavior** subscale included four items, and scores could range from zero to eight. The Cronbach Alpha is 0.69 for the 4-year-old cohort and 0.67 for the 3-year-old cohort.

[44] For the following social-emotional outcomes at the 3rd grade level, a lower score indicates better functioning and a negative impact is desirable: total child behavior problems, aggressive behavior, hyperactive behavior, withdrawn behavior, and conflict with teacher.

- The **Hyperactive Behavior** subscale included three items, and scores could range from zero to six. The Cronbach Alpha is 0.65 for the 4-year-old cohort and 0.69 for the 3-year-old cohort.

- The **Withdrawn Behavior** subscale included three items, and scores could range from zero to six. The Cronbach Alpha is 0.56 for the 4-year-old cohort and 0.60 for the 3-year-old cohort.

- **Teacher-Child Relationship.** Teachers were asked to rate their relationship with students in the study using a five-point response format ("definitely does not apply," "not really," "neutral or not sure," "applies sometimes," or "definitely applies"). The scale included items such as, "This child and I always seem to be struggling with each other," "If upset, this child will seek comfort from me," and "This child values his/her relationship with me." The 15-item instrument, developed by Robert Pianta (1996), generates three dimensions: closeness, conflict, and positive relationships.

  - The **Closeness with Teacher** dimension, focusing on positive effect, consisted of seven items, such as, "It is easy to be in tune with what this child is feeling," and "When I praise this child, he/she beams with pride." Scores can range from seven to 35 and the Cronbach Alpha for both age cohorts is 0.82.

  - The **Conflict with Teacher** dimension is measured by eight items that indicate the level to which the teacher and student are at odds with each other, such as, "This child easily becomes angry at me," and "When this child is in a bad mood, I know we're in for a long and difficult day." Scores can range from eight to 40. The Cronbach Alpha is 0.89 for the 4-year-old cohort and 0.90 for the 3-year-old cohort.

  - The **Positive Teacher-Child Relationship** dimension is a measure of the overall relationship between the student and the teacher by combining the closeness and conflict items (reversed). The score for the 15 items can range from 15 to 75. The Cronbach Alpha is 0.87 for the 4-year-old cohort and 0.88 for the 3-year-old cohort.

- **Strengths and Difficulties Questionnaire.** Teachers were asked to rate children on both problem and pro-social behavior using a three-point response format (not true, somewhat true, or certainly true). The scores for this measure were reported as the percentage of children in the normal category on the following scales.

  - **Emotional Symptoms Scale.** This scale included behavior items such as often worries, easily scared, nervous or clingy in new situations and complains of headaches or stomach aches. The five-item scale has a Cronbach Alpha of 0.72 for both age cohorts.

  - **Conduct Problems Scale.** This scale included behavior items such as loses temper, fights or bullies other children, lies or cheats, and steals from home and elsewhere. The five-item scale has a Cronbach Alpha of 0.80 for both age cohorts.

- **Hyperactivity Scale.** This scale included behavior items such as restless, constantly fidgeting, easily distracted and poor attention span. The five-item scale has a Cronbach Alpha of 0.86 for both age cohorts.

- **Peer Problem Scale.** This scale included items such as prefers to play alone, not generally liked by other children, and gets along better with adults than other children. The five-item scale has a Cronbach Alpha of 0.61 for both age cohorts.

- **Pro-social Scale.** This scale included behavior items such as whether the child is considerate of others, shares readily, is helpful if someone is hurt or sick is kind, and offers to help others. The five-item scale has a Cronbach Alpha of 0.86 for both age cohorts.

- **Total Difficulties Score.** This scale was created by summing the scores from all the scales except for the Pro-social Scale. The 20-item scale has a Cronbach Alpha of 0.87 for both age cohorts.

- **Social Competency.**[45] Teachers were asked to rate a child's emotional self-regulation and pro-social behavior using a five-point response format (not at all, a little, moderately well, well, or very well). The scale included items such as "Expresses needs and feelings appropriately" and "Can calm down when excited or all wound up." The nine-item IRT scale has a Cronbach Alpha of 0.93 for both age cohorts.

- **Children's Self-Report of Their Academic and Social Skills.** This instrument[46] was used with 3rd grade children in the ECLS-K to provide a self-assessment of their academic and social skills. The instrument used a four-point response format (not at all true, a little bit true, mostly true, or very true). The measure included four scales:

  - **Internalizing.** This scale included items about the extent to which children feel sad, lonely, or anxious such as feeling "sad a lot of the time," feeling lonely, feeling ashamed of mistakes, and worrying about school and friendships. The eight-item IRT scale has a Cronbach Alpha of 0.74 for both age cohorts.

  - **Externalizing.** This scale included items about problems with anger or distractibility problems such as fighting and arguing "with other kids," talking and disturbing others, and problems with paying attention. The six-item IRT scale has a Cronbach Alpha of 0.72 for both age cohorts.

[45] Items taken from the Social Competency Scale—Teacher Version (Fast Track Project, University Park, PA: Pennsylvania State University, 1990).

[46] The instrument was adapted from the *Self-Description Questionnaire-1* (Marsh, 1990).

- **Peer Relations.** This scale included items about how easily they make friends and get along with children as well as their perception of their popularity. The six-item IRT scale has a Cronbach Alpha of 0.72 for both age cohorts.

- **School.** This scale included items about how well they do in "all school subjects" and their enjoyment of "all school subjects." The six-item IRT scale has a Cronbach Alpha of 0.70 for both age cohorts.

Health Domain[47]

Health measures are based on parent report and included the following measures.

- **Receipt of Health Care Services.** Parents were asked to report on whether the child had received or had access to, two health care services.

  - **Whether the Child Has Health Insurance.** Parents were asked if the child is currently 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 seen a dentist since September.

- **Child's Health Status**. Parents were asked to report on their child's current 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. A dichotomous variable was developed—those who reported that their child's health is excellent or very good and those who reported that their child's health is good, fair, or poor.

  - **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 is coded yes if the parent reported any such occurrences in the last month.

[47] For all health outcomes at the 3rd grade level, a higher score indicates better functioning.

Parenting Practices Domain[48]

Parenting practices measures were based on both parent and teacher report and included the following measures.

- **Spanking.** 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?" In previous parent interviews, a question was included on use of timeout, however, this item was dropped for third grade due to the lack of clarity in the interpretation of the data.

- **Parenting Styles.** The parents were asked to respond to selected items from the Child-Rearing Practices Report (CRPR) (Block, 1965). Parents were asked to respond to items, such as, "I do not allow my child to get angry with me" and "I am easygoing and relaxed with my child" using a Likert scale that ranged from one (exactly like you) to five (not at all like you). The parenting styles were derived to assess the degree to which the two dimensions of warmth (nurturing and supportive behaviors) and control (disciplinary strategies, restrictive behaviors) exist in the parent-child relationship, based on the parents' reports. This scale has a Cronbach Alpha of 0.73 for the 4-year-old cohort and 0.71 for the 3-year-old cohort. The parenting styles identified for the analysis include:

  - **Authoritative** -- parents were identified as being high in both warmth and control,

  - **Permissive** -- parents were high in warmth but low in control,

  - **Authoritarian** -- parents exhibited high levels of control but low levels of warmth, and

  - **Neglectful** -- parents were low in both warmth and control.

- **Parent Report of the School as a Supportive Environment.** Parents were asked to respond to a list of items using a five-point response category ranging from strongly disagree to strongly agree. The items addressed school safety, as well as how well the school meets the child's academic, social, and behavioral needs, and the parent's confidence in the school. The scale included items such as "The staff at your child's school is doing good things for your child," "Your child's school is safe," and "Your child's school meets (his/her) academic needs." The eight-item IRT scale has a Cronbach Alpha of 0.89 for both age cohorts.

- **Effect of Parenting on Parent's Life.** Parents were asked to respond to a series of items related to their life as a parent using a five-point response format ranging from strongly disagree to strongly agree. The scale included items such as "I find myself

[48] For the following parenting practices outcomes at the 3rd grade level, a lower score indicates better functioning and a negative impact is desirable: spanking, and authoritarian, neglectful, and permissive parenting styles.

giving up more of my life to meet my children's needs than I ever expected," "Since having a child, I feel that I am almost never able to do things that I like to do," and "Sometimes I feel my child doesn't like me and doesn't want to be close to me." The IRT developed 12-item scale has a Cronbach Alpha of 0.89 for both age cohorts.

- **Parent Perception of School Services.** Parents were asked to respond to a series of items on their communication with the school using a three-point response format (does it very well, just OK, or does not do it at all). The scale included items such as how well the school "Lets you know (between report cards) how your child is doing in school," "Provides workshops, materials, or advice about how to help children learn at home," and "Understands the needs of families who don't speak English." The six-item IRT scale has a Cronbach Alpha of 0.77 for both age cohorts.

- **Doing Things Together.** Parents were asked a series of questions about family activities using a three-point response format (never or hardly ever, sometimes, or often). The scale included items such as "Do you and your child do things together at home?" and "Does your child go with members of the family to movies, sports events, or other outings?" The seven-item IRT scale has a Cronbach Alpha of 0.63 for both age cohorts.

- **Time Spent With Child.** Parents were asked questions about the amount of time the parent and child were together on weekdays and weekends and when with their child the amount of time they spent on an activity together. The response format is less than 30 minutes, 30 minutes to one hour, one to three hours, three to five hours, or more than five hours. The four-item IRT scale has a Cronbach Alpha of 0.67 for both age cohorts.

- **Teacher Report of Parent Participation in School Activities.** Teachers were asked two questions: "Have one or both of the child's parents (or guardians) attended open house meetings, back-to-school nights, or class events, such as a class play or recital, this year?" and "Have one or both of this child's parents (or guardians) acted as volunteers or helped out with class activities or class trips this year?"

- **Teacher Report of School Contact/Communication.** Teachers were asked two questions: "How often has this child's parents (or guardians) initiated contact with you to find out how things were going with the child or to offer help with class activities?" and "How often have you had to contact or tried to contact this child's parent(s) or guardians about behavior or schoolwork problems this child has been having?" The response categories ranged from zero (not at all) to four (about once a month or more often).

Third Grade School Characteristics

The 3rd grade school experiences measures included school measures, classroom measures, and teacher measures. Most of the **school measures** were reported by the principal with one student-reported item. School measures fall into four categories: school type and size,

educational needs and challenges of the student population, availability and quality of instructional resources, and use of data for school decisions.

- **School Type and Size** measures included: (1) type of school attended (i.e., public or private), (2) total school enrollment during the child's 3rd grade year, and (3) whether or not the school's enrollment exceeds the school's capacity.

- **Educational Needs and Challenges** were described using an IRT scale and nine individual items.

  - **Instructional Challenges.** Principals were asked to respond to a list of items using a three-point response category (not a factor, somewhat a factor, or serious factor). The items addressed issues that principals may consider as factors that prevent a school from improving. The scale included items such as "Lack of community support," "Faculty apathy and resistance to change," and "State or federal mandates." The 19-item IRT scale has a Cronbach Alpha of 0.85 for both age cohorts.

  - In addition to this scale, this construct included nine individual items assessing: (1) the student turnover within the school year (i.e., leavers and new enrollees as a percentage of the fall enrollment), (2) percentage of students who receive free or reduced price lunch, (3) percentage of non-white students, (4) percentage of Dual Language Learners (DLL), (5) percentage of students with individualized education programs (IEPs), (6) school in need of improvement status[49] (no/yes), (7) school's Title I status (no/yes), (8) school discipline problems as measured by the amount of time the principal spends on discipline issues (less than daily/daily), and (9) school safety as measured by the amount of time the student reports that she/he feels safe at school (less than daily/daily).

- **Availability and Quality of Instructional Resources** were described using two IRT scales and 13 individual items.

  - **Adequacy of School Facilities.** Principals were asked to respond to a list of items using a five-point response category (do not have, never adequate, often not adequate, sometimes adequate, or always adequate). The items addressed the availability of a cafeteria, computer lab, library/media center, art room, gymnasium, music room, playground, classrooms, auditorium, multipurpose room, science room/lab, and health room. The 12-item IRT scale has a Cronbach Alpha of 0.79 for both age cohorts.

  - **Technology at School.** Principals were asked to respond to a list of items using a four-point response category (not a barrier, small barrier, moderate barrier, or great barrier). The items addressed issues that principals may

[49] School improvement status as defined by the Elementary and Secondary Education Act (ESEA) as amended by No Child Left Behind (NCLB), 2001.

consider barriers to teachers' use of computers or the internet for instruction. The scale included items such as "Not enough computers," "Lack of appropriate professional development for teachers," and "Difficulty recruiting and hiring computer teachers or technology coordinators." The six-item IRT scale has a Cronbach Alpha of 0.79 for both age cohorts.

- In addition to these two scales, there were 13 individual items that examined: availability of special programs, including (1) gifted and talented programs, (2) before- and after-school programs, and (3) summer school (yes/no), (4) availability of instructional specialists including reading specialists and (5) Dual Language Learner (DLL) specialists (no/yes), (6) class size as measured by the regular classroom teacher-child ratio, (7) staff mobility or turnover as measured by the number of new teacher hires, (8) technology at school as measured by the ratio of children to computers, (9) percentage of instructional computers connected to the internet, (10) the availability of a technology coordinator, and (11) the availability of a technology coordinator to develop curriculum to help teachers integrate technology into classroom lessons (no/yes), (12) instructional time as measured by the average number of minutes available daily for core subject instruction as reported by the school principal (excludes lunch, recess, physical education, art, music, and other non-core subjects in a school), and (13) ability grouping as measured by the use of regrouping by ability for instruction within grade (no/yes).

- **Use of Data for School Decisions** measures to what extent the principal and his/her leadership team used data to determine school priorities or make school decisions. Specifically, principals were asked whether they use (1) standardized test scores, other formal assessments, or academic grades and (2) direct classroom observations or surveys of students, parents, or teachers to determine school priorities or make school decisions (no/yes). Principals also were asked to if they use standardized test results "to a great extent" to (1) set goals for individual student achievement, (2) compare subgroups of students (i.e., gender, race), (3) examine trends in student performance over time, and (4) examine trends in teacher performance over time.

The teacher-reported **classroom measures** fall into three categories: resources, math instruction, and reading/language arts instruction.

- **Resource** measures were described using two IRT scales and three individual items.

  - **Technology for Instruction**. Teachers were asked to respond to a list of items using a five-point response format (never, once or twice a semester, once or twice a month, once or twice a week, or daily or almost daily). The items addressed how often teachers asks students to use computers for selected activities. The scale included items such as "Practice drills," "Analyze or graph data," and "Do research using the internet." The IRT developed six-item scale has a Cronbach Alpha of 0.77 for both age cohorts.

- **Parent Supports.** Teachers were asked to respond to a list of items using a five-point response format ranging from strongly disagree to strongly agree. The items measured teacher's perceptions of their relationship with parents, communication with parents, and the extent that parents and teachers work together. The scale included items such as "The principal expects teachers to communicate regularly with parents," "Teachers really try to understand parents' concerns," and "Parents have confidence in teachers' expertise." The 12-item IRT scale has a Cronbach Alpha of 0.90 for both age cohorts.

- In addition to these two scales, there were three individual items that addressed: (1) teachers who report that their school provides all the instructional resources needed, (2) teachers who report that their classroom has at least one computer with internet access, and (3) teacher report on the percentage of time spent daily on academic instruction.

- **Math Instruction** measures were described using one IRT measure and eight individual items.

  - **Math Instructional Approach.** Teachers were asked to respond to a list of items using a variety of response formats. The items measured various components of the teachers' math instruction, including emphasis on specific math goals, high quality instruction, classroom organization, and frequency of use of selected math activities. The scale included items such as "How much emphasis do you give to understanding the concepts behind mathematics?" (none, little, moderate, or major), "During the last year, how often would you say the lessons were focused on studying a topic in depth, rather than covering the basic facts or concepts?" (none, a few, about half, or most), "How much time during a typical day do students spend on working one-on-one with the teacher?" (no time, little, time, some time, or most of the time), and "How often do children explain how a math problem is solved?" (never, once or twice a month, two or three times a month, once or twice a week, three or four times a week, or every day). The 35-item IRT scale has a Cronbach Alpha of 0.89 for both age cohorts.

  - In addition to this scale, math teachers answered eight questions regarding: (1) the average number of minutes spent on math instruction per week for the study child's class, (2) use of ability grouping for the study child' math class, (3) the number of children with serious reading difficulties in the study child's math class, (4) the number of DLL children in the study child's math class, (5) the number of disruptive children in the study child's math class, (6) the number of children below grade level in math in the study child's math class, (7) the number of children in the study child's math class, and (8) the availability of teaching assistants for the study child's math class (no/yes).

- **Reading/Language Arts Instruction** measures were described using one IRT measure and eight individual items.

  - **Reading/Language Arts Instructional Approach.** Teachers were asked to respond to a list of items using a variety of response formats. The items measured various components of the teachers' reading/language arts instruction, including having students work on various writing activities, frequency of use of comprehension activities as a focus on instruction, classroom organization, and frequency of use of selected reading/language arts activities. The scale included items such as "How often did children revise their own work by elaborating or extending what they wrote?" (never, rarely, sometimes, or often), "How often was focusing on summarizing important or critical details a focus of your instruction?" (never, rarely, sometimes, or often), "How much time during a typical day do students spend on working one-on-one with teacher?" (no time, little, time, some time, or most of the time), and "How often do you or someone else discuss new or difficult vocabulary?" (never, once or twice a month, two or three times a month, once or twice a week, three or four times a week, or every day). The 48-item IRT scale has a Cronbach Alpha of 0.91 for both age cohorts.

  - In addition to the single scale, reading/language arts teachers answered eight individual items related to (1) the average number of minutes spent on reading/ language arts instruction per week for the study child's class, (2) the use of ability grouping for the study child' reading/language arts class, (3) the number of children with serious reading difficulties in the study child's reading/ language arts class, (4) the number of DLL children in the study child's reading/language arts class, (5) the number of disruptive children in the study child's reading/language arts class, (6) the number of children below grade level in reading/language arts in the study child's reading/language arts class, (7) the number of children in the study child's reading/language arts class, and (8) the availability of teaching assistants for the study child's reading/language arts class (no/yes).

School measures reported by both teachers and principals fall into three categories: school supports, teacher skills and attitudes, and teacher qualifications.

- **School Supports** measures were described using four IRT measures.

  - **School Leadership.** Teachers were asked to rate their principal on a number of factors, including communicates school goals, establishes positive relationships, effectively manages the school, and understands student learning using a five-point response category ranging from strongly disagree to strongly agree. The scale included items such as "The principal at this school communicates a clear vision for this school" and "The principal at this school knows what is going on in my classroom." The 15-item IRT scale has a Cronbach Alpha of 0.95 for both age cohorts.

- **Instruction.** Teachers were asked to respond to questions about instructional programs in the school and commitment to these programs using a five-point scale ranging from strongly disagree to strongly agree. The scale included items such as "Once we start a new instructional program, we follow up to make sure it is working" and "You can see real continuity from one instructional program to another at this school." The six-item IRT scale has a reliability of 0.79 for both cohorts.

- **Teacher Collaboration.** Teachers were asked questions in two areas—the extent to which they agree that different types of teacher collaboration are happening in their school and the amount of time spent working with faculty in their school. The scale included items such as "To what extent do you agree or disagree that teachers share and discuss student work with other teachers?" (strongly disagree, disagree, neither agree or disagree, agree, or strongly agree) and "This school year, how often did you work with other faculty on clarifying standards for student learning through analysis of students' work?" (less than once per month, two to three times per month, once or twice a week, or daily). The 11-item IRT scale has a Cronbach Alpha of 0.89 for both age cohorts.

- **Training Opportunities.** Teachers were asked to respond to a list of items on how often they have participated in professional development or school planning activities. The scale included items such as "How much professional development or other training have you received this year in using student assessments to tailor instruction for students' individual needs?" (none, a little, some, or a great deal), and "How often have you participated in a network of teachers outside this school?" (never, once or twice, three to six times, or more than six times). The 16-item IRT scale has a Cronbach Alpha of 0.87 for both age cohorts.

- **Teacher Skills and Attitudes** measures were described using three IRT measures. These three scales were principal-reported.

  - **School Level Teacher Ratings.** Principals were asked to provide a rating of the school's faculty, the school's learning environment, and teachers' beliefs and values related to the mission of the school using a five-point response format ranging from strongly disagree to strongly agree. The scale included items such as "Teachers have a 'can do' attitude" and "Teachers in this school have a good grasp of the subject matter they teach." The 14-item IRT scale has a Cronbach Alpha of 0.80 for both age cohorts.

  - **Math Teacher Ratings.** Principals were asked to rate teachers of study children on math classroom teaching techniques and classroom instructional strategies using a five-point response format ranging from substantially below average to substantially above average. The scale included items such as "Uses advance planning to meet student learning needs in math" and "Uses a variety of materials, techniques, and resources to actively engage students in

curriculum activities during math." The 12-item IRT scale has a Cronbach Alpha of 0.93 for both age cohorts.

- **Reading/Language Arts Teacher Ratings.** This scale is similar to the math teacher ratings scale. Principals were asked to rate teachers of study children on reading/language arts classroom teaching techniques and classroom instructional strategies using a five-point response format ranging from substantially below average to substantially above average. The scale included items such as "Uses advance planning to meet student learning needs in reading/language arts" and "Uses a variety of materials, techniques, and resources to actively engage students in curriculum activities during reading/language arts." The 12-item IRT scale has a Cronbach Alpha of 0.93 for both age cohorts.

- **Teacher Attitudes.** Teachers were asked to respond to a list of items about how well teachers work together in the school to promote student learning, their satisfaction with aspects of the school, and their control over student learning. The scale included items such as "How many teachers in this school feel responsible that all students learn?" (none, some, about half, most or nearly all), "How satisfied are you with student motivation to learn?" (very dissatisfied, somewhat dissatisfied, somewhat satisfied, or very satisfied), and "How much can you do to motivate students who show no interest in their work?" (nothing, very little, some, quite a bit, a great deal). The 33-item IRT scale has a Cronbach Alpha of 0.94 for both age cohorts.

- **Teacher Qualifications** measures were described using five individual items.

  - Teachers answered five items related to: (1) the extent to which teachers agree that instruction should meet individual student needs, (2) teacher's highest degree, (3) teacher's major field of study, (4) type of teacher certification or credential, and (5) number of years of teaching experience.

Defining and Balancing the Analysis Sample

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 teachers are weighted and analyzed according to the children they describe. This makes all impact findings representative of all newly entering Head Start children in the nation in 2002 in communities in which there were more potential program participants than funded Federal Head Start slots and in programs that meet the study criteria for eligibility.

The annual cross-sectional samples are chosen to maximize the data available for analysis each spring. Thus, they include every completed child assessment, parent interview, or

teacher-child report for that year. In each instance, the comparability of the treatment and control group samples established at random assignment is maintained to the greatest extent possible by adjusting the initial sampling weights to offset observable differences between respondents and non-respondents for each cohort.

A comparison of the characteristics of the children and families in the 3rd grade impact analysis sample, using characteristics measured at baseline in fall 2002, is presented in Appendix B for both age cohorts. The comparisons use baseline data on demographic characteristics and baseline data on variables measured as outcomes in later years (i.e., pretest measures). Tables are organized to present the comparisons by the respondent source, i.e., for the direct child assessment outcomes, for parent-reported outcomes, and for outcomes reported by teachers and principals. In these tables, observations are weighted to reflect the share of the national population they represent, i.e., the 3rd grade impact analysis sample seeks to replicate this population as accurately as possible even though not every child can be included. Any differences between the program and control groups in these tables reflect the combined consequences of (1) chance differences between the treatment and control groups created at random assignment, (2) differential nonresponse in the 3rd grade data collection following weighting adjustment, and (3) for pretest measures, the possible early impacts of Head Start in fall 2002 before baseline data could be collected.

Adjusting for differences between the treatment and control group analysis samples by including the fall 2002 measures in the 3rd grade impact regressions will offset two of these factors. In particular it will reduce the threat of nonresponse bias and increase the statistical precision of the impact findings by offsetting chance differences present at baseline that continue to influence 3rd grade outcomes. Unfortunately, it will also cause the 3rd grade estimates to omit the earliest impacts of Head Start on participating children if any of those occurred prior to collection of included fall 2002 measures. Most of the fall 2002 data were collected during a three-month period from October 2002 through December 2002 (with most collected by mid-November) at a lag from the start of Head Start participation in the treatment group in August and September 2002. If impacts of Head Start occurred quickly that fall, inclusion of the fall measures as covariates in the 3rd grade impact regressions will attenuate the impact estimate—i.e., bias it toward zero—since the portion of the impact achieved prior to fall 2002 data collection would be removed from the 3rd grade findings.

In our judgment, attenuation of the 3rd grade impact findings through removing early impacts of Head Start from estimates by including pretest variables would do more harm than excluding pretest variables, thereby failing to adjust for chance differences at random assignment or any nonresponse differences that remain in the outcome data following reweighting.[50] Reflective of this perspective, we include pretest measures as covariates in the impact regressions only if strong evidence exists that an early impact of Head Start *did not occur* prior to collection of the candidate pretest in fall 2002.

Data Analysis

All of the analyses reported in the following chapters are based on comparisons of the average outcome level for the treatment group with the average outcome level for the control group, using weighted data.[51] These simple weighted difference-in-means impact estimates are reported, along with statistical tests showing which of the measured outcome differences are unlikely to be the result of chance treatment-control differences and hence can appropriately be interpreted as an effect of the Head Start program.

To add to the explanatory power of the impact analysis and further adjust for differences in the baseline characteristics of the 3rd grade analysis samples for the treatment and control groups due to outcome data collection nonresponse, key demographic variables measured in fall 2002 were included as covariates in regression models used to estimate the impact of Head Start, if they were available (these are shown in Exhibit 2.8). The selected demographic variables met two criteria: (1) they were expected to correlate with child and family outcomes (and therefore to help to increase the explanatory power of the model and reduce nonresponse bias), and (2) they could not have been influenced by Head Start during the first weeks of participation (i.e., prior to the time they were measured).

---

[50] Holding off from adjusting for chance differences between the treatment and control groups at baseline affects only the statistical precision, not the biasedness or unbiasedness of the 3rd grade impact estimates. Failing to make adjustment for differences created by nonresponse in follow-up data collection does affect bias, but not in a known direction. In contrast, attenuation due to removal of early impacts through inclusion of baseline covariates will consistently bias impact estimates toward zero, consistently understating the impact of the Head Start program.

[51] The weights account for the different rates at which primary sampling units (PSU's), grantee/delegate agencies, centers, and children within centers were sampled and at which children of different types are missing from the 3rd grade follow-up data. Thus, the study sample can be used to accurately represent the national Head Start population of all children served in eligible "non-saturated" communities.

Exhibit 2.8. Baseline Demographic Variables Included in the Statistical Models Estimating the Impact of Head Start[52]

Child Covariates
Child Gender
Child Age at Time of Spring Assessment
Child Race/Ethnicity (White/Other, Black, Hispanic)
Child's Primary Language at Baseline (English vs. Spanish/Other)
Number of Weeks Elapsed between 9/1/2002, and Date of Testing (for child assessment outcomes)
Parent Covariates
Primary Language Spoken at Home (English vs. Spanish/Other)
Primary Caregiver's Age as of 9/1/2002
Both Biological Parents Live with Child
Biological Mother Is a Recent Immigrant
Mother's Highest Level of Educational Attainment (Less Than High School, High School, Beyond High School)
Mother's Marital Status (Not Married, Married, Separated/Divorced/Widowed)
Mother Gave Birth to Study Child as a Teenager (i.e., 18 Years-0ld or Younger)
Number of Weeks Elapsed between 9/1/2002, and Parent Interview (for all other outcomes derived from the parent interview)

For all statistically significant results, the impact estimates in their initial units are converted into effect sizes by dividing by the measured standard deviation of the outcome in the population. This provides a yardstick for gauging the quantitative importance of the estimated impact in relation to the natural variation of the outcome. Effect sizes show the size of any impact relative to the distribution of outcomes that would have prevailed had no Head Start intervention been available.

The analysis of main impacts generated a large number of individual statistical tests. Such conditions increase the probability that one or more statistically significant differences will emerge by random chance alone in the absence of a true impact—an event known as a "false discovery." To guard against such false discoveries, Benjamini and Hochberg (1995) developed a statistical test designed to screen out marginally significant findings from large sets of impact

[52] Missing values on the baseline demographic variables, due to both item and instrument nonresponse in fall 2002, were imputed using hot deck imputation, a procedure in which observations with missing values are filled in with values taken from children with similar characteristics on the other baseline variables. To ensure comparability, the distribution of each variable was compared before and after imputation to check that the imputation procedures had not appreciably changed the dataset.

estimates. This procedure was applied to the complete set of outcomes within each domain (cognitive, social-emotional, health, and parenting outcomes). This was done separately for each of the two study cohorts. Because the Benjamini-Hochberg test limits discovery of true impacts compared with conventional test procedures, we present findings both with and without the Benjamini-Hochberg procedure. We have used these tests to establish three separate categories of statistically significant results[53] and use this language throughout this report:

- **Strong Evidence:** the estimated impact for a particular outcome is statistically significant in a conventional hypothesis test ($p \leq 0.05$), and this result holds up after adjusting for false discovery.
- **Moderate Evidence:** the estimated impact for a particular outcome is statistically significant in a conventional hypothesis test ($p \leq 0.05$), but this result *does not* hold up after adjusting for false discovery.
- **Suggestive Evidence:** the estimated impact for a particular outcome is statistically significant only when a relaxed standard ($p \leq 0.10$) is used in a conventional hypothesis test. This result *may or may not* hold up after adjusting for false discovery.

Estimating Variations in Impact

In addition to looking at Head Start's impact *on average*, it is important to understand how impact *varies* among different types of participants. Congress mandated that the study identify circumstances under which the program achieves its greatest impact, in terms of both child and family circumstances—what works best for which children? There is also interest in determining whether the benefits of Head Start measured for children and families in general are widespread—i.e., whether gains compared to the control group occur for many types of children and families, or whether the overall gain reflects big gains for some participants and little or no gain for others.

There are many examples of variations in program impact that would have important policy or program implications. Analyses can spotlight groups of children who are not advanced by their participation and suggest needs for program improvement. For example, Head Start

[53] If we fail to identify a statistically significant difference, we do not have conclusive evidence that the program "doesn't work." Rather, statistically insignificant impacts mean that the effect is indeterminate—access to Head Start may or may not have had a non-zero impact on a particular outcome, and we cannot with this study sample make a confident conclusion either way. The one thing that will be known with confidence is that a large true impact has not occurred.

programs are required to serve children with special needs, making it important to understand the extent to which these children benefit from their participation.

To examine which children benefit, and from which kind of family circumstances at program entry, the study analyzed the impact of Head Start on subgroups of program participants, defined by child and family characteristics at baseline. Some typical questions to be addressed in this realm include: Does Head Start help Dual Language Learners? Does it help children with the lowest pre-academic skills at baseline?

A computationally efficient and statistically powerful way to examine such subgroup impacts uses an extension of the impact regression models discussed above. Interaction terms are added between the subgroup-defining variables and the indicator of Head Start assignment. The coefficients on the interactions show how impacts vary between subgroups. Impacts for the individual subgroups can be obtained from the regression, as can differences in impact between any two subgroups. For example, for subgroups defined by a mother's race/ethnicity, a single regression can provide information on how large an impact Head Start had on children of White mothers, children of Black mothers, and children of Hispanic mothers, as well as how impacts vary across these subgroups. Separate regressions are run for each subgroup-defining factor, referred to in this report as a subgroup "dimension," such as special needs status, household risk index, and urban/rural location. Each analysis tests for variation in impact along a particular dimension, using the entire sample of children available in a given age cohort—an improvement in terms of statistical precision over separate analysis of each subgroup one at a time.

Exhibit 2.9 lists the dimensions used to define subgroups for analysis. These dimensions were chosen in advance of conducting the analyses, based on (1) their program and policy importance to the Administration for Children and Families, (2) past Head Start and child development research, and (3) recommendations from members of the Advisory Committee on Head Start Research and Evaluation. Impacts are examined for each subgroup, separately by age cohort, to determine how widespread the benefits of Head Start might be and to identify the child/family types and the program characteristics associated with larger impacts. Details of subgroup dimensions used, their rationales, and the subgroup impact estimation method appear in Chapter 5.

Exhibit 2.9. Factors Used To Define Subgroups

Child and Family Characteristics*
Special Needs (Yes/No)
Child's Pre-Academic Skills 2002 Baseline as Assessed on the Woodcock-Johnson III Pre-Academic Composite Measure (Yes/No)
Child's Home Language (English, Dual Language Learner)
Biological Mother/Caregiver Race/Ethnicity (White, Black, Hispanic)
Household Risk Index (Low/No, Medium, High)
Urbanicity (Urban, Non-Urban)
Parent/Caregiver-Reported Depressive Symptoms (No, Mild, Moderate, or Severe Depressive Symptoms)

*These subgroups are defined in Chapter 5.

Impacts on Head Start Participants

The main impact estimates in this report measure the effect of Head Start on the average child randomly assigned to the Head Start treatment group—that is, the impact of *granting access to Head Start services* for the population randomized. However, as discussed earlier in this chapter, not all of the children given access to Head Start in the study sites actually participated in federally funded Head Start services, the intended treatment. As a result, there are two different versions of the key research question of this 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 typical child and family who *actually participated* in Head Start, on average?

In addition to children given access to Head Start but not attending (known as no-shows), some of the families of children randomized into the *control group* managed to get their children into Head Start anyway. This subpopulation is known as "crossovers." The Head Start Impact Study had no way to fully ensure that the children and families randomly assigned to the control group did not participate in federally funded Head Start. A total of 17.6 percent of the weighted sample in the control group are known to have participated in a federally funded Head Start program for some time during the first year of the study. The presence of no-shows and crossovers changes the meaning of the experimental comparison between the full treatment

group and the full control group; it becomes the impact of *intent to treat* (ITT). The impact of actual *receipt* of the Head Start intervention (compared to non-receipt)—remains important to policy. This leads to interest in estimates of the "*impact on the treated*" (IOT), which show how Head Start affects the outcomes of a set of children who participate in Head Start compared to what would have happened to those same children had they not participated.

While methods to adjust for no-shows are fairly straightforward, adjusting for crossovers is more challenging. After assessing multiple research options we concluded that the best way to provide information on Head Start's IOT impact is through the use of an "instrumental variable" (IV) methodology. This methodology uses assignment to the treatment group as a statistical instrument for participation in the program. This method, long known in the econometric literature and applied in recent years to random assignment evaluations of government social programs,[54] treats crossovers symmetrically with no-shows and adjusts the initial ITT impact estimates to remove the influence of both groups in attenuating the magnitude of the estimates. This is achieved by dividing the ITT estimate by 1 minus the no-show rate minus the crossover rate to get an "impact on the treated" (IOT) impact estimate: IOT = ITT / (1 – n – c), where n is the no-show rate, and c is the crossover rate.

Like the classic "Bloom no-show adjustment" (Bloom, 1984), this methodology postulates that a program's overall impact on the treatment group accrues to just a subset of the sample. The Bloom assumption is that no impact occurs for no-shows since they are never exposed to the intervention; this is widely viewed as an innocuous assumption. The IV methodology further assumes that Head Start's impact on crossovers equals, on average, the program's impact on the corresponding children in the treatment group—i.e., the children who would have crossed over and participated in Head Start had they been assigned to the control group. The combination of these two assumptions makes it appropriate to reallocate the total impact of the program observed in the contrast between the full treatment and control groups to just the remaining set of children whose impacts are neither zero nor offsetting: essentially, the children who comply with the intention of random assignment by participating in Head Start when randomized into the treatment group and not participating when randomized into the

[54] See, for example, L.A. Gennetian, P.A. Morris, J.M. Bos, and H.S. Bloom (2005). Constructing instrumental variables from experimental data to explore how treatments produce effects. *Learning more from social experiments: Evolving analytic approaches*. H. S. Bloom (Ed.), New York: Russell Sage Foundation.

control group. This average impact on "compliers" can be inferred from the ITT impact estimate without further analysis, since the ITT estimate under the assumptions above is just the average impact on the "compliers" and a zero net impact on everyone else.

It should be noted, however, that for the IOT measure defined here to characterize *all* Head Start participants, it must be true that crossover-equivalent children in the treatment group experience the same impact on average as other Head Start participants in the treatment group.[55] Hence, the key assumptions in the crossover adjustment are:

- equal impact for "crossover-like" children in the treatment group and other Head Start participants in the treatment group; and
- equal outcomes on average for "crossovers" in the control group and "crossover-like" children in the treatment group.

These assumptions are discussed in detail in the Technical Report for the Head Start Impact Study which provides the reasoning and empirical evidence from the field to justify them in the Head Start context. As explained there, tests of statistical significance for the original ITT estimates apply as well to the IOT. That is, adjusting for no-shows and crossovers changes the magnitude of the estimated impacts, but not their statistical significance.

Understanding Impacts on the 3-Year-Old Cohort[56]

As noted earlier, the control group for the 3-year-old cohort was given access to Head Start, by study design, in the second year of the study. Excluding children from Head Start for two years was considered both infeasible and undesirable because it likely would undermine Head Start programs' willingness to participate in the study. Therefore, the design does not allow the study to ask what is the impact of two years in Head Start or what is the differential impact of one versus two years of Head Start. However the design allows the study to determine whether having Head Start available at age three is helpful to children brought to the program at that age,

---

[55] This is necessary so that IOT = ITT/ (1 – n – c), which relates directly to just the "compliers" in the participant population, is a good estimate of the effect on all other participants as well—i.e., on the children in the treatment group who would be crossovers if assigned to the control group but who as part of the treatment group comprise part of the total set of all participants in the treatment group, the group of policy interest in IOT analysis.

[56] In a sense, one can think of the 4-year-old cohort as "delay-entry" eligible 3-year-olds, i.e., parents who had the option to enroll their child at age three but opted instead to 'hold them back" until age four. Not surprisingly, as discussed elsewhere, the parents who elected to seek enrollment for their child at age three are different from those who decided to wait until age four, and our data bear this out. For example, the 3-year-old cohort of newly entering children is more African American and the 4-year-old cohort of newly entering children is more Hispanic.

or whether those children would be just as well off, initially and over the longer term, if the program did not enroll them until age four. This is not only important to individual families, it also answers an important policy question. To answer the question, the study precluded program entry at age three for the control group while allowing it at age four for those children, then contrasted outcomes after that point with statistically equivalent children never excluded from the program (i.e., treatment group). Therefore, the research design for the 3-year-old cohort only varied the first year of Head Start participation. Hence, impacts for the 3-year-old cohort reflect the benefits of being provided an *earlier* year of Head Start given that a later, age 4 year will be available.

This difference has important implications for how IOT estimates are calculated. Control group members who attended Head Start in that second year are not considered crossovers, since they did not violate random assignment. Therefore, they are not part of the IV adjustment when computing IOT impact estimates in spring 2004 and beyond. It is only members of the 3-year-old control group who attended Head Start during that first year who are included in these IOT adjustments.

Chapter 3: School Experiences

Introduction

Exposure to Head Start can potentially yield effects on children's elementary school experiences by improving parents' ability to advocate for their children's elementary school education. Possible pathways that parents can choose to affect their child's school, classroom or teacher characteristics are:

- **Taking Advantage of School Choice Options** – in recent years parents nationally have had increased opportunities to select schools for their children. These choices can include not only the historically available options of private school education (for low-income parents this has generally involved parochial schools), but increasingly also choices among alternative local public elementary schools (called "open enrollment"), public charter schools, magnet schools, and publicly funded school vouchers.

- **Advocating for Better Access to Resources and More Effective Teachers in School** – once children are enrolled in a school, parents can be more engaged in their child's education by advocating for placement with particular teachers, or for opportunities to access within-school educational resources (e.g., enrichment activities, tutoring).

In both cases, Head Start's potential impact on parents may lead to comparatively better early educational experiences that may, in turn, have subsequent impacts on children's cognitive and/or social-emotional outcomes. In this chapter findings are presented on the differences in school experiences for the Head Start group children and the control group children.

Measuring 3rd Grade School Experiences

As discussed in Chapter 2, we collected information on the nature and quality of children's 3rd grade educational experiences from multiple sources including teachers and school principals. The results reported in this chapter focus on the characteristics of the schools children attended and the outcomes presented are primarily based on data reported by school principals and the classroom teachers of the study children.

The measures fall into three broad categories: overall characteristics of the child's school, instructional characteristics of the child's classroom, and the characteristics of children's math and reading/language arts teachers. The measures are reported as follows:

- **School Characteristics:** to assess Head Start's impact on the types of elementary schools children attend, we examined impacts on the following school-level measures:

  - **School Type and Size:** affiliation (public vs. private),[57] enrollment size, and possible overcrowding.

  - **Educational Needs and Challenges:** student mobility, student body characteristics (poverty, ethnicity, Dual Language Learners, students with disabilities), student discipline problems, student-reported safety concerns, and principal-reported perceptions of school-level instructional challenges.

  - **Instructional Resources:** adequacy of school facilities, availability and quality of educational technology, availability of special programs (e.g., reading specialists, tutors, gifted and talented programs), average classroom teacher-student ratio, staff turnover, average daily instructional time, and use of instructional grouping.

  - **Use of Data for Educational Decisions:** what types of data are used (e.g., test scores, classroom observations) and for what types of school- or student-level decisions.

- **Classroom Characteristics:** to assess Head Start's impact on the quality of the classes children attend in 3rd grade, we examined the following classroom-level measures:

  - **Available Classroom Resources:** access to, and use of, technology for instruction, teacher communication with parents and parent-teacher relationship, teacher-reported adequacy of instructional resources, and**,** the percent of time spent on academic instruction daily.

  - **Mathematics Instruction:** instructional goals for math, classroom organization for math instruction, frequency of use of particular math instructional strategies, daily math instructional time, instructional grouping, availability and use of teaching assistants, and the characteristics of the students in the study child's class for math instruction (e.g. percentage of children in the class with discipline problems).

[57] Note: the data collected did not include information on whether the child's school was a charter or magnet school.

- **Reading/Language Arts Instruction:** instructional goals for reading/language arts, classroom organization for reading/language arts instruction, frequency of use of particular reading/language arts instructional strategies, daily reading/language arts instructional time, instructional grouping, availability and use of teaching assistants, and the characteristics of the students in the study child's class for reading/language arts instruction.

- **Teacher Characteristics:** to assess impacts of Head Start on the types of teachers children have in 3rd grade, we assessed the characteristics and views of the children's teachers:

  - **Available School Supports for Teachers:** teacher perceptions of the extent of administrator support for teachers, commitment to consistent instructional programs and goals, opportunities for teacher collaboration, and opportunities for teacher professional development.

  - **Teacher Skills and Attitudes:** principal-reported ratings of study children's teachers related to general classroom management and specific math and reading/language arts instruction, and teacher-reported satisfaction with their school, their teacher colleagues, and their ability to affect their student's academic learning.

  - **Teacher qualifications:** educational qualifications, years of teaching experience, and teaching certification.

Results

The two age cohorts of children tended to end up in many schools. By the end of 3rd grade, the study sample had dispersed into nearly 1,600 individual elementary schools. About 39 percent of the schools enrolled only students from the 3-year-old cohort, about 45 percent enrolled only students from the 4-year-old cohort, and 16 percent of the schools enrolled students from both age cohorts.

Overall, there were few school experiences differences at the end of 3rd grade. Results are provided for both age cohorts at the end of this chapter in Exhibit 3.1 for the 3-year-old cohort and in Exhibit 3.2 for the 4-year-old cohort. As discussed in Chapter 2, the following language categorizes results:

- **Strong Evidence*:*** the estimated impact is statistically significant at the 5 percent level ($p \leq 0.05$), and the result holds up under the test for multiple comparisons.

- **Moderate Evidence:** the estimated impact is statistically significant at the 5 percent level ($p \leq 0.05$), but the result *does not* hold up under the test for multiple comparisons.

- **Suggestive Evidence:** the estimated impact is statistically significant under a relaxed standard ($p \leq 0.10$), and the result *may or may not* hold up under the test for multiple comparisons.

3-Year-Old Cohort

As shown in Exhibit 3.1, there were a limited number of statistically significant impacts on school-level measures of children's educational experiences for the 3-year-old cohort. There was moderate evidence that principals rated school facilities as more adequate and that there was less staff turnover at schools attended by Head Start group children.

There was also suggestive evidence that the Head Start children were more likely than the non-Head Start children to attend public, rather than private schools and to attend schools where a higher percentage of 3rd grade students score at the proficient or higher level on the state reading/language arts assessment. This is despite suggestive evidence that the schools attended by Head Start children had a higher percentage of students with disabilities and moderate evidence that these schools required more attention by the principal to deal with discipline problems.

In terms of classroom measures, there was moderate evidence that the Head Start children were more likely to have access to an instructional aide in their reading/language arts classroom and suggestive evidence that they had a lower percentage of children reading below grade level in their reading/language arts class. On teacher measures, Head Start children were more likely to have teachers who majored in education as undergraduates.

4-Year-Old Cohort

As shown in Exhibit 3.2, unlike the 3-year-old cohort, there were no statistically significant impacts on the type of school attended by Head Start children in the 4-year-old cohort, nor on the average state test scores for 3rd grade students in their school.

There were mixed findings on other school-level measures, although the majority of them favor the Head Start children. There was moderate evidence that Head Start children were

more likely to attend a school that offers summer school and had greater access to computers than control group children. There also was suggestive evidence that Head Start children attended schools that were more likely to use standardized tests to compare subgroups of children. Finally, there was suggestive evidence of higher mobility for students in the schools that Head Start children attended compared to the control group.

There were no significant impacts on the classroom measures for the 4-year-old cohort. In terms of teacher measures, there was suggestive evidence that the Head Start children were more likely to have a teacher with a standard state teaching certificate.

Summary

This chapter examined the potential effect of exposure to Head Start on the school experiences of children at the end of 3rd grade. Possible effects were assessed across a range of school, classroom, and teacher characteristics. For the most part, there were few differences between Head Start and non-Head Start children. Those differences that were found were modest in size and included some that are favorable and some that are unfavorable.

For children in the 3-year-old cohort, Head Start children were in schools that had more adequate school facilities as reported by principals, lower staff turnover, and a higher percentage of 3rd grade students scoring at the proficient or higher level on the state reading/language arts assessment. This is despite the fact that the schools attended by Head Start children had higher percentages of students with disabilities and according to principals, required more attention to deal with student discipline problems. In terms of classroom and teacher measures, the Head Start children were more likely to have access to an instructional aide in their classroom, to have a lower percentage of children reading below grade level in their reading/ language arts class, and to have a teacher who majored in education as an undergraduate.

For children in the 4-year-old cohort, there were no statistically significant impacts on the type of school attended or on the average state test scores for 3rd grade students in their school. There was, however, evidence that the Head Start children were in schools that, according to principals, had greater access to computers (compared to the non-Head Start children) and were more likely to have summer school programs. Head Start children were more likely to have a teacher with a standard state teaching certificate and their schools were more

likely to use standardized tests to a great extent to compare subgroups of students. On the other hand, the schools attended by the Head Start children were more likely to have higher levels of student mobility, as reported by principals.

Exhibit 3.1. Estimated Impacts on School Experiences at 3rd Grade: 3-Year-Old Cohort

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
3rd Grade Year (Spring 2008)							
SCHOOL MEASURES							
School Type and Size							
Percent in Public Schools	97.00	95.00	2.00	0.112	3.00*	0.068	0.14
Total School Enrollment, 2008	546.82	546.42	0.39	0.981	1.9	0.896	0.01
Percent in Overcrowded Schools	77.00	79.00	-2.00	0.548	-3.00	0.474	-0.07
Educational Needs and Challenges							
Adequacy of Facilities++	0.06	-0.11	0.17	0.018	0.21**	0.012	0.21
Percent of Student Turnover in Schools Attended by Study Children	15.00	15.00	1.00	0.561	1.00	0.567	0.05
Percent of Students Receiving Free/ Reduced Price Lunch	70.13	70.02	0.12	0.924	-0.94	0.558	-0.04
Percent of non-White Students	58.06	60.11	-2.05	0.342	-3.24	0.118	-0.09
Percent of DLL Students	13.79	11.76	2.03	0.146	2.23	0.101	0.12
Percent of Students with IEP	13.14	12.45	0.69	0.309	1.34*	0.053	0.14
Percent in Schools Identified as in Need of Improvement	79.00	77.00	2.00	0.453	1.00	0.820	0.02
Percent in Title I Schools	91.00	89.00	2.00	0.360	1.00	0.578	0.04
Percent in Schools Where Principals Spend Time on Discipline At Least Daily	31.00	24.00	7.00	0.027	8.00**	0.029	0.19
Percent of Students Who Feel Safe at School Every Day	42.00	40.00	2.00	0.764	0.00	0.930	-0.01

Exhibit 3.1. Estimated Impacts On School Experiences At 3rd Grade: 3-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Average Percent of Students in 3rd Grade at or Above Proficient in Math	67.04	66.61	0.43	0.723	1.93	0.121	0.10
Average Percent of Students in 3rd Grade at or Above Proficient in Reading/LA	66.32	64.21	2.11	0.062	2.42*	0.058	0.10
Availability and Quality of Instructional Resources							
Instructional Challenges++	-0.03	0.06	-0.01	0.201	-0.11	0.128	-0.11
Technology at School++	0.06	-0.01	0.06	0.293	0.08	0.287	0.08
Percent in Schools with Gifted and Talented Programs	69.00	66.00	3.00	0.187	3.00	0.434	0.05
Percent in Schools with Before and After School Programs	87.00	88.00	-1.00	0.575	-2.00	0.321	-0.07
Percent in Schools with Summer School	80.00	79.00	1.00	0.620	0.00	0.904	-0.01
Percent in Schools with Reading Specialists	77.00	77.00	0.00	0.938	1.00	0.617	0.04
Percent in Schools with ELL Specialists	52.00	51.00	1.00	0.645	2.00	0.581	0.03
Number of Students Per Regular Classroom Teacher	20.40	19.96	0.44	0.184	0.43	0.275	0.07
Percent of Staff Turnover in Schools Attended by Study Children	12.00	13.00	-1.00	0.154	-2.00**	0.036	-0.14
Number of Students Per Computer	9.25	10.08	-0.83	0.654	-0.6	0.733	-0.02

Exhibit 3.1. Estimated Impacts On School Experiences At 3rd Grade: 3-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Percent of Instructional Computers Connected to Internet	95.23	94.10	1.13	0.270	0.91	0.479	0.05
Percent in Schools with a Technology Coordinator	61.00	62.00	-1.00	0.708	1.00	0.636	0.03
Percent in Schools with a Technology Coordinator Who Helps Integrate Technology into Lessons	47.00	51.00	-4.00	0.241	-2.00	0.487	-0.05
Average Daily Minutes Spent on Core Subjects	318.10	316.20	1.90	0.528	-1.03	0.760	-0.02
Percent in Schools Using Ability Grouping in 3rd Grade	73.00	72.00	1.00	0.778	-4.00	0.310	-0.08
Use of Data for School Decisions							
Percent in Schools Using Test Scores, Assessments, and Grades for Deciding School Priorities	88.00	88.00	0.00	0.986	-1.00	0.501	-0.04
Percent in Schools Using Classroom Observation and Surveys for School Priorities	63.00	64.00	-1.00	0.844	-2.00	0.523	-0.05
Percent in Schools Using Standardized Tests to a Great Extent to Set Goals for Individual Student Achievement	66.00	68.00	-2.00	0.601	-3.00	0.354	-0.07

Exhibit 3.1. Estimated Impacts On School Experiences At 3rd Grade: 3-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Percent in Schools Using Standardized Tests to a Great Extent to Compare Subgroups of Students	50.00	54.00	-3.00	0.330	-6.00	0.210	-0.11
Percent in Schools Using Standardized Tests to a Great Extent to Examine Trends in Student Performance Over Time	66.00	71.00	-5.00	0.133	-5.00	0.176	-0.11
Percent in Schools Using Standardized Tests to a Great Extent to Examine Trends in Teacher Performance Over Time	47.00	48.00	-1.00	0.637	-2.00	0.531	-0.05
CLASSROOM MEASURES							
Resources							
Technology for Instruction++	0.13	-0.01	0.15	0.024	0.08	0.226	0.09
Parent Supports++	0.02	-0.04	0.06	0.401	0.07	0.325	0.07
Percent in Schools with Teachers Who Report Having All Needed Instructional Resources	21.00	18.00	2.00	0.502	2.00	0.528	0.05
Percent in Schools With Internet Access in Classrooms	86.00	87.00	-1.00	0.683	-1.00	0.734	-0.03
Overall Daily Instructional Time	76.85	75.72	1.13	0.415	1.26	0.353	0.08
Math Instruction							
Math Instructional Approach++	0.08	0.02	0.06	0.223	0.05	0.378	0.06
Length of Math Instruction in Minutes Per Week	348.46	346.32	2.14	0.732	-1.67	0.82	-0.02

Exhibit 3.1. Estimated Impacts On School Experiences At 3rd Grade: 3-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Percent in Math Classes Using Ability Grouping	24.00	21.00	2.00	0.381	1.00	0.856	0.01
Percent of Students with Serious Reading Difficulties in Math Class	20.00	21.00	-1.00	0.259	-1.00	0.378	-0.06
Percent of DLL Students in Math Class	10.00	11.00	-1.00	0.586	-1.00	0.720	-0.03
Percent of Disruptive Students in Math Class	11.00	11.00	0.00	0.972	0.00	0.837	0.02
Percent of Students Below Grade in Math Class	24.00	25.00	-1.00	0.348	-2.00	0.164	-0.10
Average Number of Students per Math Class	17.31	17.06	0.25	0.576	0.14	0.758	0.02
Percent in Math Classes with a Teaching Assistant	27.00	24.00	2.00	0.456	2.00	0.469	0.06
Reading/Language Arts Instruction							
Reading Instructional Approach++	0.01	-0.02	0.03	0.703	0.02	0.844	0.02
Length of Reading/ LA Instruction in Minutes Per Week	550.57	537.53	13.05	0.231	15.07	0.147	0.08
Percent in Reading/ LA Classes Using Ability Grouping	43.00	43.00	0.00	1.00	0.00	0.959	0.00
Percent of Students with Serious Reading Difficulties in Reading/LA Class	20.00	21.00	-1.00	0.307	-2.00	0.203	-0.09
Percent of DLL Students in Reading/ LA Class	11.00	11.00	0.00	0.814	0.00	0.834	-0.01

Exhibit 3.1. Estimated Impacts On School Experiences At 3rd Grade: 3-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Percent of Disruptive Students in Reading/ LA Class	11.00	11.00	-1.00	0.641	0.00	0.833	-0.02
Percent of Students Below Grade in Reading/LA Class	29.00	31.00	-3.00	0.131	-3.00*	0.067	-0.13
Average Number of Students per Reading/LA Class	15.86	15.83	0.04	0.936	-0.05	0.917	-0.01
Percent in Reading/ LA Classes with a Teaching Assistant	36.00	29.00	6.00	0.010	7.00**	0.014	0.15
TEACHER MEASURES							
School Supports							
School Leadership++	0.02	0.04	-0.02	0.799	0.01	0.883	0.01
Instruction++	0.03	-0.01	0.04	0.502	0.05	0.528	0.05
Teacher Collaboration++	0.08	0.01	0.08	0.275	0.11	0.138	0.11
Training Opportunities++	0.04	-0.03	0.07	0.255	0.10	0.143	0.10
Teacher Skills and Attitudes							
School Level Teacher Ratings++	0.02	0.03	0.00	0.948	-0.03	0.605	-0.03
Math Teacher Ratings++	0.07	0.03	0.03	0.699	0.04	0.671	0.04
Reading Teacher Ratings++	0.04	0.06	-0.02	0.749	-0.03	0.756	-0.03
Teacher Attitudes++	0.02	-0.03	0.05	0.567	0.06	0.428	0.06
Teacher Qualifications							
Percent of Students with Teachers Who Agree or Strongly Agree That Instruction Should Meet Individual Student Needs	93.00	92.00	1.00	0.571	1.00	0.684	0.03

Exhibit 3.1. Estimated Impacts On School Experiences At 3rd Grade: 3-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Percent of Students with Teachers Who Have Education Beyond BA	53.00	50.00	3.00	0.401	3.00	0.424	0.07
Percent of Students with Teachers Who Have a Major in Education	86.00	82.00	4.00	0.267	7.00*	0.051	0.18
Percent of Students with Teachers Who Have Standard Certification	92.00	90.00	2.00	0.256	2.00	0.417	0.06
Average Years of Teaching Experience	12.65	12.95	-0.29	0.652	0.06	0.933	0.01

Key:

*** $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

++ Indicates that the scores are reported as IRT scores with a mean of 0 and a standard deviation of 1. IRT scores are reported rather than true scores because a true score is a non-linear transformation of an IRT score, which can change the distribution and variance of the construct.

There are no significant outcomes using the Benjamini-Hochberg test for multiple comparisons.

Exhibit 3.2. Estimated Impacts on School Experiences at 3rd Grade: 4-Year-Old Cohort

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
3rd Grade Year (Spring 2007)							
SCHOOL MEASURES							
School Type and Size							
Percent in Public Schools	98.00	98.00	0.00	0.828	0.00	0.830	-0.02
Total School Enrollment, 2007	573.04	568.59	4.46	0.784	-8.18	0.650	-0.03
Percent in Overcrowded Schools	70.00	71.00	-1.00	0.685	-2.00	0.620	-0.04
Educational Needs and Challenges							
Adequacy of Facilities++	0.01	-0.09	0.11	0.122	0.12	0.157	0.13
Percent of Student Turnover in Schools Attended by Study Children	14.00	14.00	0.00	0.644	1.00*	0.084	0.10
Percent of Students Receiving Free/ Reduced Price Lunch	68.29	66.32	1.97	0.408	1.13	0.607	0.04
Percent of non-White Students	59.78	59.29	0.49	0.850	0.37	0.869	0.01
Percent of DLL Students	22.35	24.94	-2.58	0.173	-1.31	0.461	-0.04
Percent of Students with IEP	10.96	11.13	-0.17	0.749	0.03	0.957	0.00
Percent in Schools Identified as in Need of Improvement	83.00	83.00	0.00	0.907	-1.00	0.669	-0.04
Percent in Title I Schools	88.00	85.00	3.00	0.348	3.00	0.399	0.08
Percent in Schools Where Principals Spend Time on Discipline At Least Daily	27.00	27.00	0.00	0.995	0.00	0.976	0.00
Percent of Students Who Feel Safe at School Every Day	38.00	44.00	-6.00	0.300	-4.00	0.437	-0.09

Exhibit 3.2. Estimated Impacts on School Experiences at 3rd Grade: 4-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Average Percent of Students in 3rd Grade at or Above Proficient in Math	65.31	65.19	0.12	0.900	-1.04	0.478	-0.06
Average Percent of Students in 3rd Grade at or Above Proficient in Reading/LA	59.53	59.62	-0.09	0.957	-1.51	0.369	-0.06
Availability and Quality of Instructional Resources							
Instructional Challenges++	-0.02	-0.08	0.06	0.484	0.10	0.258	0.10
Technology at School++	0.03	0.06	0.03	0.672	0.00	0.971	0.00
Percent in Schools with Gifted and Talented Programs	74.00	72.00	2.00	0.552	3.00	0.351	0.07
Percent in Schools with Before and After School Programs	91.00	84.00	7.00	0.156	7.00	0.119	0.20
Percent in Schools with Summer School	81.00	72.00	9.00	0.009	11.00***	0.004	0.25
Percent in Schools with Reading Specialists	78.00	82.00	-4.00	0.167	-4.00	0.177	-0.11
Percent in Schools with ELL Specialists	58.00	57.00	1.00	0.774	2.00	0.524	0.04
Number of Students Per Regular Classroom Teacher	21.25	21.43	-0.18	0.684	-0.04	0.933	-0.01
Percent of Staff Turnover in Schools Attended by Study Children	13.00	13.00	0.00	0.787	0.00	1.00	0.00
Number of Students Per Computer	8.98	13.62	-4.64	0.114	-6.23**	0.050	-0.15

Exhibit 3.2. Estimated Impacts on School Experiences at 3rd Grade: 4-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Percent of Instructional Computers Connected to Internet	93.47	94.65	-1.18	0.466	-0.96	0.594	-0.07
Percent in Schools with a Technology Coordinator	68.00	65.00	2.00	0.505	2.00	0.487	0.05
Percent in Schools with a Technology Coordinator Who Helps Integrate Technology into Lessons	49.00	45.00	4.00	0.267	3.00	0.422	0.06
Average Daily Minutes Spent on Core Subjects	312.94	312.72	0.22	0.958	0.85	0.833	0.02
Percent in Schools Using Ability Grouping in 3rd Grade	64.00	66.00	-2.00	0.529	-3.00	0.407	-0.07
Use of Data for School Decisions							
Percent in Schools Using Test Scores, Assessments, and Grades for Deciding School Priorities	89.00	89.00	0.00	0.955	1.00	0.800	0.03
Percent in Schools Using Classroom Observation and Surveys for School Priorities	61.00	61.00	1.00	0.893	0.00	0.960	0.00
Percent in Schools Using Standardized Tests to a Great Extent to Set Goals for Individual Student Achievement	63.00	68.00	-5.00	0.170	-6.00	0.117	-0.13

Exhibit 3.2. Estimated Impacts on School Experiences at 3rd Grade: 4-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Percent in Schools Using Standardized Tests to a Great Extent to Compare Subgroups of Students	54.00	48.00	6.00	0.160	7.00*	0.053	0.14
Percent of Schools Using Standardized Tests to a Great Extent to Examine Trends in Student Performance Over Time	68.00	65.00	3.00	0.340	2.00	0.600	0.04
Percent of Schools Using Standardized Tests to a Great Extent to Examine Trends in Teacher Performance Over Time	50.00	45.00	5.00	0.148	2.00	0.548	0.04
CLASSROOM MEASURES							
Resources							
Technology for Instruction++	-0.15	-0.23	0.08	0.213	0.08	0.230	0.08
Parent Supports++	-0.01	-0.08	0.07	0.357	0.09	0.174	0.10
Percent in Schools with Teachers Who Report Having All Needed Instructional Resources	17.00	20.00	-4.00	0.254	-3.00	0.387	-0.07
Percent in Schools with Internet Access in Classrooms	86.00	84.00	2.00	0.422	2.00	0.371	0.06
Overall Daily Instructional Time	77.42	77.73	-0.31	0.755	-0.015	0.887	-0.01
Math Instruction							
Math Instructional Approach++	-0.15	-0.14	-0.01	0.875	0.00	0.990	0.00
Length of Math Instruction Per Week	336.81	330.21	6.60	0.312	3.51	0.593	0.04

Exhibit 3.2. Estimated Impacts on School Experiences at 3rd Grade: 4-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Percent in Math Classes Using Ability Grouping	21.00	20.00	1.00	0.712	0.00	0.909	-0.01
Percent of Students with Serious Reading Difficulties in Math Class	20.00	18.00	2.00	0.320	1.00	0.411	0.09
Percent of DLL Students in Math Class	17.00	19.00	-1.00	0.502	-1.00	0.531	-0.05
Percent of Disruptive Students in Math Class	11.00	10.00	0.00	0.715	0.00	0.928	0.01
Percent of Students Below Grade in Math Class	26.00	23.00	3.00	0.115	2.00	0.254	0.12
Average Number of Students per Math Class	18.39	18.53	-0.14	0.777	-0.43	0.390	-0.08
Percent in Math Classes with a Teaching Assistant	19.00	19.00	0.00	0.879	2.00	0.421	0.06
Reading/Language Arts Instruction							
Reading Instructional Approach++	-0.08	-0.00	-0.07	0.243	-0.01	0.109	-0.11
Length of Reading/ LA Instruction in Minutes Per Week	585.21	582.83	2.38	0.889	-8.45	0.629	-0.04
Percent in Reading/ LA Classes Using Ability Grouping	47.00	47.00	0.00	0.927	0.00	0.937	0.01
Percent of Students with Serious Reading Difficulties in Reading/LA Class	20.00	19.00	0.00	0.876	1.00	0.658	0.03
Percent of DLL Students in Reading/ LA Class	16.00	19.00	-3.00	0.259	-1.00	0.508	-0.04

Exhibit 3.2. Estimated Impacts on School Experiences at 3rd Grade: 4-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Percent of Disruptive Students in Reading/ LA Class	11.00	11.00	0.00	0.924	-1.00	0.561	-0.04
Percent of Students Below Grade in Reading/LA Class	31.00	30.00	1.00	0.740	0.00	0.803	0.02
Average Number of Students per Reading/LA Class	16.94	17.03	-0.09	0.876	-0.28	0.651	-0.05
Percent in Reading/ LA Classes with a Teaching Assistant	31.00	29.00	2.00	0.498	3.00	0.418	0.07
TEACHER MEASURES							
School Supports							
School Leadership++	-0.06	-0.04	-0.02	0.794	0.01	0.849	0.01
Instruction++	-0.08	-0.06	-0.01	0.902	0.01	0.891	0.01
Teacher Collaboration++	-0.09	-0.17	0.08	0.347	0.12	0.130	0.11
Training Opportunities++	-0.05	-0.03	-0.02	0.800	0.00	0.995	0.00
Teacher Skills and Attitudes							
School Level Teacher Ratings++	0.01	-0.07	0.07	0.323	0.09	0.259	0.08
Math Teacher Ratings++	-0.01	-0.11	0.09	0.228	0.07	0.445	0.08
Reading Teacher Ratings++	-0.09	-0.08	0.00	0.967	-0.05	0.643	-0.05
Teacher Attitudes++	-0.08	-0.09	0.01	0.853	0.02	0.717	0.02
Teacher Qualifications							
Percent of Students with Teachers Who Agree or Strongly Agree That Instruction Should Meet Individual Student Needs	91.00	92.00	-1.00	0.725	-1.00	0.809	-0.02

Exhibit 3.2. Estimated Impacts on School Experiences at 3rd Grade: 4-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Percent of Students with Teachers Who Have Education Beyond BA	50.00	50.00	0.00	0.970	-1.00	0.880	-0.01
Percent of Students with Teachers Who Have a Major in Education	82.00	84.00	-1.00	0.555	-1.00	0.587	-0.03
Percent of Students with Teachers Who Have Standard Certification	94.00	90.00	4.00	0.040	3.00*	0.075	0.11
Average Years of Teaching Experience	13.62	13.20	0.42	0.542	0.53	0.471	0.05

Key:

*** $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

++ Indicates that the scores are reported as IRT scores with a mean of 0 and a standard deviation of 1. IRT scores are reported rather than true scores because a true score is a non-linear transformation of an IRT score, which can change the distribution and variance of the construct.

There are no significant outcomes using the Benjamini-Hochberg test for multiple comparisons.

Chapter 4: The Impact of Head Start on Child and Parent Outcomes at the End of 3rd Grade

Introduction

This chapter focuses on the overall average impact of Head Start on child and parent outcomes. It examines the impact of access to Head Start and the impact of participation in Head Start. Chapter 5 presents findings for subgroups of children. As described in Chapter 2, the study measured outcomes through direct child assessments, parent and child surveys, and reports from children's teachers for both the 3- and 4-year old cohorts. First, the chapter presents the intent to treat (ITT)[58] analyses (impacts of access to Head Start) for the four outcome domains: cognitive outcomes, social-emotional outcomes, health outcomes, and parenting practices outcomes. These sections also compare outcomes for the study sample with national averages when these are available. Second, the chapter presents impacts on the treated (IOT), showing the impacts of participation in Head Start. Finally, this chapter discusses the 3rd grade findings in light of the previous Head Start Impact Study findings (Head Start/pre-K through 1st grade) and provides a summary of the findings. Appendix C presents detailed tables on the baseline characteristics for the 3rd grade analysis sample, and the estimated ITT impacts at each spring follow-up through 3rd grade (2003-2008). Appendix D presents the estimated 2003-2008 IOT impacts for these spring follow-ups.

As discussed in Chapter 2, the following language categorizes results:

- **Strong Evidence*:*** the estimated impact is statistically significant at the 5 percent level ($p \leq 0.05$), and the result holds up under the test for multiple comparisons.
- **Moderate Evidence*:*** the estimated impact is statistically significant at the 5 percent level ($p \leq 0.05$), but the result *does not* hold up under the test for multiple comparisons.
- **Suggestive Evidence*:*** the estimated impact is statistically significant under a relaxed standard ($p \leq 0.10$), and the result *may or may not* hold up under the test for multiple comparisons.

The chapter, in general, does not provide narrative descriptions of results that do not meet these criteria, though they are presented in tables.

[58] See Chapter 2 for a discussion of intent to treat (ITT) and impact on the treated (IOT) estimates.

Intent to Treat (ITT) Impacts

3rd Grade Impacts on Cognitive Outcomes

The cognitive domain consisted of: (1) direct assessments of language and literacy skills, pre-writing skills (in Head Start years only), and math skills; (2) teacher reports of children's school performance; and (3) parent report of grade promotion. There was suggestive evidence of one impact for each age cohort. As shown in Exhibit 4.1, the Head Start group children in the 4-year-old cohort demonstrated significantly higher test scores than the non-Head Start group children on reading at the end of 3rd grade as measured by the ECLS-K (Early Childhood Longitudinal Study-Kindergarten Cohort) Reading Assessment. For the 3-year-old cohort, there was also one significant impact, though it was unfavorable--the parents of the Head Start group children reported a significantly lower child grade promotion rate than the parents of the non-Head Start group children (Exhibit 4.2).

The finding for the Spanish language and literacy measure, Batería Woodcock-Muñoz Identificación de letras y palabras, showed no significant differences between children in the Head Start and non-Head Start groups. This was the only test used to measure ongoing growth in Spanish language and literacy skills for Dual Language Learners (DLLs) who were initially tested in Spanish at baseline and then switched to English for later assessments.[59]

In addition to impacts, the study showed how the skills of this sample of low-income children, both those who received access to Head Start and those who didn't, compared to 3rd graders nationally. On national norms, both the Head Start and control groups continued to lag behind in cognitive outcomes, as is typical for low-income children. For example, on average, the HSIS children (both treatment and control) performed about one-half of a standard deviation lower than the nationally representative ECLS-K sample at 3rd grade on the Reading Assessment, or about eight points lower for the HSIS children.[60] This holds true for both age cohorts. However, as noted above for the 4-year-old cohort, Head Start group children demonstrated

[59] The Spanish-English group was identified as Dual Language Learners (DLLs). At baseline, the appropriate language of assessment for this group of children was determined to be Spanish, so the DLLs were administered a Spanish assessment with two English subtests. At subsequent data collections, the DLLs were administered an English assessment with one or two Spanish subtests. The DLL group does not include children in Puerto Rico, who were tested in Spanish at all data collection points.

[60] This ETS analysis (Najarian &Yan, 2008) was reported in the HSIS scoring report for the ECLS-K Reading Assessment.

Exhibit 4.1. Estimated Impacts on 3rd Grade Cognitive Outcomes: 4-Year-Old Cohort

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
3rd Grade Year (Spring 2007)							
Language and Literacy Measures							
ECLS-K Reading	98.61	96.63	1.98	0.139	2.23*	0.075	0.11
PPVT (Adapted)	408.14	405.74	2.40	0.298	2.17	0.246	0.08
WJ III Letter-Word Identification	482.10	480.60	1.51	0.450	2.11	0.275	0.07
Spanish Language and Literacy Measures+++							
WM Letter-Word Identification	464.78	462.31	2.47	0.787	3.53	0.678	0.07
Math Skills Measures							
WJ III Applied Problems	486.96	487.70	-0.74	0.601	-0.43	0.729	-0.02
WJ III Calculation	491.28	491.52	-0.24	0.826	0.00	0.997	0.00
School Performance Assessment Measures							
Promotion	0.96	0.95	0.00	0.885	0.01	0.768	0.02
Language and Literacy Ability	0.66	0.63	0.03	0.547	0.04	0.389	0.09
Math Ability	0.69	0.72	-0.03	0.454	-0.03	0.462	-0.07
Reading/Language Arts Skills++	-0.06	-0.05	-0.02	0.855	-0.01	0.945	-0.01
Math Skills++	-0.02	-0.01	-0.01	0.902	-0.03	0.632	-0.03

Key:

*** $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

++ These scores are reported in the Item Response Theory (IRT) metric. Since there is not a normalizing sample to link the scores to, the scores are reported in a standardized format with a mean of 0 and a standard deviation of 1.

+++ Indicates scores for only the DLLs on the mainland.

Exhibit 4.2. Estimated Impacts on 3rd Grade Cognitive Outcomes: 3-Year-Old Cohort

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
3rd Grade Year (Spring 2008)							
Language and Literacy Measures							
ECLS-K Reading	98.10	97.91	0.20	0.868	-0.18	0.876	-0.01
PPVT (Adapted)	407.85	405.67	2.18	0.122	1.83	0.146	0.06
WJ III Letter-Word Identification	483.60	482.81	0.79	0.661	0.44	0.818	0.01
Spanish Language and Literacy Measures+++							
WM Letter-Word Identification	468.63	470.77	-2.14	0.734	-1.63	0.804	-0.03
Math Skills Measures							
WJ III Applied Problems	486.44	486.48	-0.05	0.975	0.03	0.985	0.00
WJ III Calculation	491.79	491.66	0.13	0.896	-0.05	0.960	0.00
School Performance Assessment Measures							
Promotion	0.94	0.95	-0.01	0.332	-0.02*	0.092	-0.11
Language and Literacy Ability	0.70	0.73	-0.03	0.434	-0.04	0.372	-0.09
Math Ability	0.74	0.74	0.00	0.920	0.01	0.680	0.03
Reading/Language Arts Skills++	0.04	0.06	-0.02	0.822	-0.05	0.515	-0.05
Math Skills++	0.06	0.02	0.04	0.453	0.01	0.861	0.01

Key:

*** $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

++ These scores are reported in the Item Response Theory (IRT) metric. Since there is not a normalizing sample to link the scores to, the scores are reported in a standardized format with a mean of 0 and a standard deviation of 1.

+++ Indicates scores for only the DLLs on the mainland.

significantly higher test scores than the non-Head Start group children on this assessment at the end of 3rd grade. On average, HSIS children in the 4-year-old cohort scored about 8 points (about one-half of a standard deviation) lower than the ECLS-K sample while the control group children scored 10 points lower than the ECLS- K sample.

The analysis of the ECLS-K reading assessment provided estimates of the likelihood that each child in the sample was proficient[61] on eight skill sets—letter recognition, beginning sounds, ending sounds, sight words, comprehension of words in context, literal inference, extrapolation, and evaluation. Compared to the ECLS-K sample, the HSIS children had comparable proficiency probability scores on lower level skills, such as letter recognition, beginning sounds, ending sounds, and sight words. That is, they were as likely to be proficient in these skills as a nationally representative group of 3rd graders. However, on higher order skills such as comprehension of words in text, literal inference, extrapolation, and evaluation, children in the HSIS sample lagged behind their 3rd grade peers in the national ECLS-K sample. For example, the average evaluation proficiency probability score for both HSIS cohorts was 0.15 while the corresponding average proficiency probability score for the ECLS-K sample was 0.26 (Najarian &Yan, 2008). This difference is equivalent to about one-half of a standard deviation.

For the 3-year-old cohort, parents of the Head Start group children reported a lower grade promotion rate than the parents of the control group children—94 percent of the Head Start group parents and 95 percent of the control group parents reported that their child was promoted to the next grade. National grade promotion rates for 3rd grade are not easily available. Many states do not report 3rd grade promotion rates and for states that do, the rates vary from state to state. Warren and Saliba (2012) used the Current Population Surveys (CPS) and the American Community Survey (ACS) to generate a predicted 3rd grade national retention rate using an age-grade delay model as a conceptual proxy for retention. For the 2008-09 school year, Warren and Saliba predicted a 3 percent grade retention rate or a 97 percent grade promotion rate at 3rd grade, a higher promotion rate than reported by either the Head Start or control group parents.

3rd Grade Impacts on Social-Emotional Outcomes

The social-emotional domain consisted of parent-reported measures during the Head Start years and reports by both parents and teachers from kindergarten through the end of 3rd grade. Outcomes included measures of children's behavior, social skills and approaches to

[61] The proficiency probability score is based on estimates of the number of correct answers that would have been expected if each child had answered all of the 154 items in the ECLS-K kindergarten, 1st grade and 3rd grade item pool. The proficiency scores also were reported in the ETS scoring report noted footnote 4.

learning, teacher-child relationships, and school adjustment. Additionally, at 3rd grade, children were asked directly about their peer relationships and school experiences.[62]

Many measures in the social-emotional domain were scored in such a way that a lower number indicated better functioning. For example, fewer behavior problems are preferable to more behavior problems. In contrast, measures of positive behaviors and skills—such as social competency—were scored such that higher scores indicated better functioning. For 14 of the social-emotional measures reported in this chapter,[63] higher scores indicate better functioning, so impacts in the positive direction indicate benefits from Head Start. For the remaining measures, (i.e., total problem behavior, aggressive behavior, hyperactive behavior, withdrawn behavior, and conflict), a lower value indicates better functioning, so a negative impact is indicative of a benefit from Head Start.

Exhibits 4.3 and 4.4 provide the social-emotional impacts and their effect sizes at the end of 3rd grade. For children in the 4-year-old cohort, there were several statistically significant impacts at the end of 3rd grade, but there were marked differences among the impacts found for ratings by teachers, parents, and children. There were two favorable impacts on parental reports of their child's behavior, providing moderate evidence of less aggressive behavior for children in the Head Start group compared to children in the non-Head Start group and suggestive evidence of fewer total problem behaviors for the Head Start group children. However, teacher reports showed unfavorable impacts: strong evidence of an unfavorable impact on the incidence of children's emotional symptoms and suggestive evidence of unfavorable impacts on closeness and having a positive relationship with the teacher. Finally, the self-reports of children in the 4-year-old cohort showed moderate evidence of an unfavorable impact on peer relations at school for children in the Head Start group compared to the control group.

For children in the 3-year-old cohort, there was only a single statistically significant social-emotional impact at the end of 3rd grade. There was moderate evidence of children in the Head Start group demonstrating better social skills and positive approaches to learning as

[62] See Chapter 2 for a description of the social-emotional outcomes.

[63] These measures include: Social Skills and Positive Approaches to Learning; Closeness with Teacher; Positive Teacher-Child Relationship; Social Competency; the six Strengths and Difficulties measures—Conduct Problems, Emotional Symptoms, Hyperactivity, Peer Problems, Pro-social Behavior, and Total Difficulties; and the four child self-report measures—Externalizing, Internalizing, Peer Relations, and School.

Exhibit 4.3. Estimated Impacts on 3rd Grade Social-Emotional Outcomes: 4-Year-Old Cohort

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
3rd Grade Year (Spring 2007)							
Parent-Reported Measures							
Aggressive Behavior	2.24	2.47	-0.23	0.073	-0.23**	0.043	-0.13
Hyperactive Behavior	1.91	1.99	-0.07	0.520	-0.08	0.435	-0.05
Withdrawn Behavior+	1.02	1.13	-0.11	0.163	-0.11	0.187	-0.09
Total Problem Behavior	5.70	6.18	-0.47	0.137	-0.50*	0.090	-0.12
Social Skills and Positive Approaches to Learning	11.95	12.11	-0.16	0.208	-0.10	0.383	-0.05
Teacher-Reported Measures							
Conduct Problems (Percent in Normal Category)	0.76	0.80	-0.04	0.155	-0.02	0.394	-0.06
Emotional Symptoms (Percent in Normal Category)	0.89	0.94	-0.05	0.005	**-0.06*** **	0.005	-0.24
Hyperactivity (Percent in Normal Category)	0.76	0.75	0.01	0.758	0.00	0.938	-0.01
Peer Problems (Percent in Normal Category)	0.83	0.89	-0.05	0.010	-0.04	0.104	-0.11
Pro-social Behavior (Percent in Normal Category)	0.74	0.78	-0.04	0.225	-0.05	0.163	-0.13
Total Difficulties (Percent in Normal Category)	0.71	0.77	-0.06	0.117	-0.06	0.140	-0.15
Closeness with Teacher	28.03	28.56	-0.53	0.077	-0.67*	0.060	-0.13
Conflict with Teacher	14.55	13.72	0.83	0.064	0.65	0.136	0.10
Positive Teacher-Child Relationships	61.05	62.41	-1.36	0.034	-1.33*	0.063	-0.14
Social Competency++	0.02	0.12	-0.10	0.186	-0.09	0.261	-0.09

Exhibit 4.3. Estimated Impacts on 3rd Grade Social-Emotional Outcomes: 4-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Child-Reported Measures							
Externalizing++	0.00	0.09	-0.09	0.249	-0.09	0.226	-0.09
Internalizing++	0.03	0.14	-0.11	0.199	-0.10	0.212	-0.10
Peer Relations++	-0.06	0.07	-0.13	0.021	-0.14**	0.020	-0.13
School++	-0.02	-0.07	0.05	0.586	0.05	0.575	0.05

Key:

******* $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

+ Indicates the reliability (Cronbach alpha) for the measure is < 0.6.

++ These scores are reported in the Item Response Theory (IRT) metric. Since there is not a normalizing sample to link the scores to, the scores are reported in a standardized format with a mean of 0 and a standard deviation of 1.

Bold regression-adjusted impact value indicates the outcome passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.

Exhibit 4.4. Estimated Impacts on 3rd Grade Social-Emotional Outcomes: 3-Year-Old Cohort

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
3rd Grade Year (Spring 2008)							
Parent-Report Measures							
Aggressive Behavior	2.42	2.38	0.04	0.752	0.04	0.703	0.02
Hyperactive Behavior	1.88	1.87	0.01	0.931	0.01	0.942	0.00
Withdrawn Behavior	1.05	0.93	0.12	0.151	0.10	0.230	0.08
Total Problem Behavior	5.94	5.77	0.17	0.600	0.15	0.604	0.03
Social Skills and Positive Approaches to Learning	12.19	11.97	0.23	0.069	0.24**	0.025	0.12
Teacher-Reported Measures							
Conduct Problems (Percent in Normal Category)	0.76	0.75	0.02	0.629	-0.01	0.759	-0.02
Emotional Symptoms (Percent in Normal Category)	0.92	0.89	0.03	0.122	0.01	0.705	0.03
Hyperactivity (Percent in Normal Category)	0.77	0.73	0.04	0.207	0.02	0.431	0.05
Peer Problems (Percent in Normal Category)	0.85	0.82	0.03	0.128	0.00	0.984	0.00
Pro-social Behavior (Percent in Normal Category)	0.74	0.75	-0.01	0.688	-0.05	0.137	-0.12
Total Difficulties (Percent in Normal Category)	0.73	0.70	0.03	0.279	0.01	0.832	0.01
Closeness with Teacher	28.08	28.13	-0.05	0.899	-0.40	0.300	-0.08
Conflict with Teacher	14.48	14.62	-0.14	0.774	0.27	0.613	0.04
Positive Teacher-Child Relationships	61.34	61.09	0.25	0.712	-0.45	0.549	-0.04
Social Competency++	0.02	-0.00	0.02	0.737	-0.05	0.427	-0.05

Exhibit 4.4. Estimated Impacts on 3rd Grade Social-Emotional Outcomes: 3-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
3rd Grade Year (Spring 2008)							
Child-Reported Measures							
Externalizing++	-0.04	-0.02	-0.02	0.763	-0.02	0.733	-0.02
Internalizing++	-0.05	-0.07	0.02	0.753	0.02	0.731	0.02
Peer Relations++	0.04	-0.05	0.08	0.230	0.08	0.227	0.09
School++	0.09	0.05	0.04	0.596	0.05	0.564	0.04

Key:

*** $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

++ These scores are reported in the Item Response Theory (IRT) metric. Since there is not a normalizing sample to link the scores to, the scores are reported in a standardized format with a mean of 0 and a standard deviation of 1.

reported by their parents, compared with the non-Head Start group. There were no other impacts on parent-reported, teacher-reported, or child-reported measures of social-emotional development for the 3-year-old cohort.

In general, for the 4-year-old cohort, parents reported significant positive social-emotional impacts, while teachers and the children themselves reported significant unfavorable impacts. Although there was strong evidence of an unfavorable impact on the incidence of children's emotional symptoms, most children in both the Head Start group (89 percent) children and the control group (94 percent) children were in the normal category for the incidence of emotional symptoms. The one favorable impact for the 3-year-old cohort at the end of 3rd grade (social skills and positive approaches to learning) was also reported for this age cohort at earlier points, specifically at the end of the age 4 year and at the end of kindergarten.

3rd Grade Impacts on Health Outcomes

This section focuses on the health domain, which is comprised of five parent reported measures that fall into two categories: (1) children's receipt of health care services and

(2) their current health status. The study included no direct collection of health data (e.g., from direct health examinations, health records, or medical provider report).

As shown in Exhibits 4.5 and 4.6, there were no statistically significant impacts on any of the measured health outcomes at the end of 3rd grade, for either age cohort. For both cohorts over 85 percent of both Head Start and control group children had received dental care in the last year, and over 85 percent had health insurance coverage. About 80 percent had excellent/good health status according to their parents. Less than 20 percent of children had need for ongoing care or had care for an injury in the last month.

Exhibit 4.5. Estimated Impacts on 3rd Grade Child Health Outcomes: 4-Year-Old Cohort

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
3rd Grade Year (Spring 2007)							
Parent-Reported Measures							
Child Received Dental Care	0.86	0.85	0.01	0.674	0.01	0.730	0.03
Child Has Health Insurance Coverage	0.88	0.87	0.01	0.648	0.00	1.000	0.00
Child's Overall Health Status Is Excellent/Good	0.80	0.79	0.02	0.478	0.01	0.547	0.03
Child Needs Ongoing Care	0.16	0.15	0.01	0.650	0.00	0.861	0.01
Child Had Care for Injury Last Month	0.12	0.11	0.01	0.724	0.01	0.815	0.02

Key:

*** $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

Exhibit 4.6. Estimated Impacts on 3rd Grade Child Health Outcomes: 3-Year-Old Cohort

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
3rd Grade Year (Spring 2008)							
Parent-Reported Measures							
Child Received Dental Care	0.87	0.86	0.01	0.681	0.00	0.979	0.00
Child Has Health Insurance Coverage	0.90	0.89	0.01	0.476	0.02	0.354	0.06
Child's Overall Health Status Is Excellent/Good	0.80	0.81	-0.01	0.714	-0.01	0.681	-0.02
Child Needs Ongoing Care	0.19	0.15	0.03	0.179	0.02	0.421	0.06
Child Had Care for Injury Last Month	0.12	0.16	-0.04	0.217	-0.04	0.216	-0.11

Key:

*** $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

3rd Grade Impacts on Parenting Practices

The parenting practices domain for this study consisted of five categories of outcomes: (1) disciplinary practices, (2) educational supports, (3) parenting styles, (4) parent participation in and communication with the school, and (5) parent and child time together. Exhibits 4.7 and 4.8 provide the impact estimates for parenting practices at the end of 3rd grade for both cohorts.

Data from the Third Grade Follow-up demonstrated a single statistically significant favorable impact in the parenting domain for each age cohort. For children in the 4-year-old cohort, there was strong evidence of a positive impact on the time that parents reported spending with their child, with significantly greater time reported for parents of Head Start group children. For children in the 3-year-old cohort, there was moderate evidence of a favorable impact on parents' reported use of the preferred authoritative parenting style (i.e., high control and high warmth).

Exhibit 4.7. Estimated Impacts on 3rd Grade Parenting Practices Outcomes: 4-Year-Old Cohort

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
3rd Grade Year (Spring 2007)							
Parent-Reported Measures							
Parent Spanked Child in Last Week	0.23	0.25	-0.02	0.595	-0.02	0.635	-0.04
Parenting Style: Authoritarian	0.16	0.13	0.02	0.333	0.02	0.494	0.05
Parenting Style: Authoritative	0.56	0.60	-0.04	0.340	-0.02	0.516	-0.05
Parenting Style: Neglectful	0.12	0.12	0.00	0.960	0.00	0.843	-0.01
Parenting Style: Permissive	0.16	0.14	0.02	0.495	0.01	0.523	0.04
Supportive School Environment++	-0.00	-0.03	0.03	0.620	0.03	0.701	0.03
Effect of Parenting on Parent's Life++	0.05	0.01	0.04	0.603	0.06	0.450	0.06
Doing Things Together++	-0.06	-0.04	-0.02	0.796	-0.02	0.786	-0.02
Time Spent with Child++	0.05	-0.19	0.24	0.003	**0.27*** **	0.001	0.27
Parent Perception of School Services++	-0.10	-0.02	-0.09	0.145	-0.09	0.175	-0.10
Teacher-Reported Measures							
School Contact and Communication	0.66	0.72	-0.05	0.058	-0.04	0.103	-0.10
Parent Participation	0.81	0.86	-0.05	0.097	-0.01	0.652	-0.04

Key:

*** $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

++ These scores are reported in the Item Response Theory (IRT) metric. Since there is not a normalizing sample to link the scores to, the scores are reported in a standardized format with a mean of 0 and a standard deviation of 1.

Bold regression-adjusted impact value indicates the outcome passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.

Exhibit 4.8. Estimated Impacts on 3rd Grade Parenting Practices Outcomes: 3-Year-Old Cohort

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
3rd Grade Year (Spring 2008)							
Parent-Reported Measures							
Parent Spanked Child in Last Week	0.25	0.28	-0.02	0.467	-0.02	0.630	-0.04
Parenting Style: Authoritarian	0.11	0.14	-0.03	0.307	-0.03	0.298	-0.08
Parenting Style: Authoritative	0.67	0.59	0.08	0.035	0.08**	0.033	0.16
Parenting Style: Neglectful	0.06	0.08	-0.02	0.403	-0.01	0.495	-0.05
Parenting Style: Permissive	0.15	0.19	-0.04	0.172	-0.04	0.160	-0.10
Supportive School Environment++	0.05	-0.00	0.05	0.378	0.03	0.672	0.03
Effect of Parenting on Parent's Life++	-0.06	-0.03	-0.03	0.531	-0.04	0.429	-0.04
Doing Things Together++	0.00	-0.07	0.08	0.208	0.08	0.160	0.09
Time Spent with Child++	-0.04	-0.01	-0.03	0.643	-0.05	0.408	-0.05
Parent Perception of School Services++	0.09	-0.03	0.12	0.084	0.11	0.142	0.11
Teacher-Reported Measures							
School Contact and Communication	0.77	0.75	0.02	0.501	0.00	0.883	-0.01
Parent Participation+	0.81	0.84	-0.03	0.203	-0.03	0.181	-0.08

Key:

*** $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

++ These scores are reported in the Item Response Theory (IRT) metric. Since there is not a normalizing sample to link the scores to, the scores are reported in a standardized format with a mean of 0 and a standard deviation of 1.

Estimated Impacts on Participants (IOT Impacts)

This section presents Head Start's impact on those children who *actually participated in the program*, for those outcomes for which a statistically significant impact of access to Head Start was found at the end of 3rd grade. Deriving estimates of the impact of

participating in Head Start (IOT estimates) from the more directly estimated impacts of access to Head Start (ITT estimates) entails scaling the estimates by a constant factor, derived for each study age cohort.[64] There is **no** change in the statistical significance of the estimates. Looking at effects on participants does not change the overall patterns found in the main analysis. For the 4-year-old cohort, this scaling factor is 1.496, while for the 3-year-old cohort, the factor is 1.433. Exhibit 4.9 provides the estimated IOT impacts for significant outcomes for the 4-year-old cohort, while Exhibit 4.10 provides the estimated IOT impacts for the 3-year-old cohort.

Thus, estimates of the impact of participating (IOT estimates) are about 50 percent larger than the estimates of the impact of access to Head Start (ITT estimates). For example, the impact of access to Head Start on the ECLS-K Reading score at the end of 3rd grade for the 4-year-old cohort is 2.23 (ITT estimate), while the impact of participating in Head Start on this outcome is 3.34 (IOT estimate) (see Exhibit 4.9).

Similarly, the impact of access to Head Start on children's social skills and positive approaches to learning at the end of 3rd grade for the 3-year-old cohort is 0.24 (ITT estimate) while the impact of participating in Head Start on children's social skills and positive approaches to learning is 0.34 (IOT estimate) (see Exhibit 4.10).

[64] The main impact estimates in this report measure the effect of Head Start on the average child randomly assigned to the Head Start group—that is, the impact of granting access to Head Start services for the population randomized. This is the impact of intent to treat (ITT). However, not all children given access to Head Start in the study sites actually participated in Head Start (no shows), and the parents of some children selected for the non-Head Start group managed to enroll their children in Head Start (crossovers). The presence of no-shows and crossovers changes the meaning of the experimental comparison between the full treatment group and full control group. This leads to interest in estimates of the impact on the treated (IOT) which shows how Head Start affects the outcomes of a set of children who participate in Head Start compared to what would have happened to those same children had none of them participated. See Chapter 2 for detailed descriptions of ITT and IOT.

Exhibit 4.9. Estimated IOT Impacts for 3rd Grade Outcomes Showing Significant ITT Impacts: 4-Year-Old Cohort

Outcome	ITT Regression Adjusted Impact	IOT		
	Impact	Impact	p-value	Effect Size
Cognitive				
Direct Child Assessment				
ECLS-K Reading	2.23	3.34*	0.075	0.16
School Performance				
No Significant Outcomes				
Social Emotional				
Parent-Reported Measures				
Aggressive Behavior	-0.23	-0.34**	0.043	-0.19
Total Problem Behavior	-0.50	-0.75*	0.090	-0.18
Teacher-Reported Measures				
Emotional Symptoms (Percent in Normal Category)	-0.06	**-0.09*** **	0.005	-0.36
Closeness	-0.67	-1.00*	0.060	-0.19
Positive Relationships	-1.33	-1.99*	0.063	-0.21
Child-Reported Measures				
Peer Relations	-0.14	-0.21**	0.020	-0.19
Health				
No Significant Outcomes				
Parenting Practices				
Parent-Reported Measures				
Time Spent With Child	0.27	**0.40*** **	0.001	0.40
Teacher-Reported Measures				
No Significant Outcomes				

NOTE: The 4-year-old cohort compliance rate is 0.66897.

Key:

*** $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

Bold IOT impact indicates the outcome passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.

Exhibit 4.10. Estimated IOT Impacts for 3rd Grade Outcomes Showing Significant ITT Impacts: 3-Year-Old Cohort

Outcome	ITT Regression Adjusted Impact	IOT		
	Impact	Impact	p-value	Effect Size
Cognitive				
Direct Child Assessment				
No Significant Outcomes				
School Performance				
Promotion	-0.02	-0.03*	0.092	-0.16
Social Emotional				
Parent-Reported Measures				
Social Skills and Positive Approaches to Learning	0.24	0.34**	0.025	0.17
Teacher-Reported Measures				
No Significant Outcomes				
Child-Reported Measures				
No Significant Outcomes				
Health				
No Significant Outcomes				
Parenting Practices				
Parent-Reported Measures				
Parenting Style: Authoritative	0.08	0.11**	0.033	0.23
Teacher-Reported Measures				
No Significant Outcomes				

NOTE: The 3-year-old cohort compliance rate is 0.69739.

Key:

*** $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

3rd Grade Findings Within the Context of Earlier Head Start Impact Study Findings

As mentioned earlier, the Third Grade Follow-up was built upon the earlier phases of the Head Start Impact Study, which followed the study children from their entry into Head Start through 1st grade. This section briefly discusses the 3rd grade findings in light of the earlier HSIS findings. Exhibits 4.11 through 4.18 summarize the ITT findings in each of the study domains

from the Head Start year through 3rd grade by age cohort. Appendix C includes complete tables for the 2002 through 2008 ITT impact estimates, showing effect sizes. Appendix D includes the 2003-2008 IOT impact estimates with effect sizes. Chapter 5 presents findings for subgroups.

Impacts on Cognitive Outcomes

The Head Start Impact Study found impacts for the sample as a whole at the end of one year of Head Start on a broad range of early language and literacy outcomes for children in both the 3- and 4-year-old cohorts, with impacts on math skills for children in the 3-year-old cohort. However, these early effects rapidly dissipated in elementary school, with only a single impact remaining at the end of 3rd grade for children in each age cohort. The data indicated that the initial Head Start benefits are quickly "made up" by children in the non-Head Start group.

This pattern is illustrated in Exhibit 4.11 by the progression of the mean scores on the Woodcock-Johnson III Letter-Word Identification test, which was measured at all data points. As shown in the exhibit, although children in both the treatment and control groups continued to increase their levels of pre-literacy achievement over time, the two groups made more substantial gains at different time points. In both age cohorts, the Head Start children made greater gains than the control group during the initial year of the study, when they were first assigned either to Head Start or to the control group. In contrast, the next year the control group children made greater gains than the Head Start group. In this year, the program and control groups experienced similar services--most of the 4-year-old cohort entered kindergarten, while among the 3-year-old cohort more similar proportions of program and control group children participated in Head Start and other pre-K programs. In addition, for the 3-year-old cohort, the control group also made greater gains than the program group in their kindergarten year. After kindergarten, the treatment and control groups in both cohorts advanced at roughly the same pace. In brief, the pattern showed initial accelerated gains for the Head Start children, then those gains were quickly made up by the control group children, followed by continued gains at the same pace for both groups. The same pattern can be demonstrated for the PPVT scores.

Exhibit 4.11. ITT Impacts on the Woodcock-Johnson III Letter-Word Identification Test, by Year and Age Cohort

Data Point	4 Year Old Cohort				3 Year Old Cohort			
	Head Start: Mean Score	Annual Gain	Non-Head Start: Mean Score	Annual Gain	Head Start: Mean Score	Annual Gain	Non-Head Start: Mean Score	Annual Gain
Baseline	310		307		295		293	
End of First Pre-K Year	325	+15	319	+12	307	+12	301	+8
End of 2nd Pre-K Year					333	+26	330	+29
Kindergarten	378	+53	378	+59	384	+51	383	+53
1st Grade	433	+55	432	+54	434	+50	433	+50
3rd Grade	482	+49	481	+49	484	+50	483	+50

NOTE: There was a significant difference on the WJ III Letter-Word Identification test between the treatment and control group for both age cohorts at the end of the first pre-K year.

For children in the 4-year-old cohort, there was strong evidence that the Head Start group demonstrated better skills at the end of the Head Start year on six outcomes related to children's language and literacy development: (1) Peabody Picture Vocabulary Test (PPVT), (2) Woodcock-Johnson III (WJ III) Letter-Word Identification, (3) WJ III Spelling, (4) WJ III Pre-Academic Skills, (5) Color Identification, and (6) Letter Naming. At the end of Head Start, parents also reported that their children had more emerging literacy skills than did parents of children in the control group. By the end of kindergarten and in later follow-ups, there were few impacts in this domain for the 4-year-old cohort. At the end of kindergarten, there were no statistically significant impacts on any of the cognitive outcomes, while at the end of 1st grade, there was suggestive evidence of a positive impact on vocabulary scores (PPVT). Likewise, at the end of 3rd grade, there was evidence of a positive impact on reading skills, where the Head Start group children demonstrated significantly higher test scores than the non-Head Start group children on reading skills as measured by the ECLS-K Reading Assessment.

For children in the 3-year-old cohort, there was strong evidence of an impact on the following six cognitive outcomes at the end of their first Head Start year: (1) PPVT (vocabulary), (2) WJ III Letter-Word Identification, (3) Preschool Comprehensive Test of Phonological and Print Processing (CTOPPP) Elision, (4) Letter Naming, (5) WJ III Pre-Academic Skills, and (6) McCarthy Draw-A-Design (prewriting). There was moderate evidence of an impact on WJ III

Applied Problems (math skills). Favorable impacts of Head Start were also found on parental reports of children's emergent literacy skills at the end of the Head Start year. In later years, the impacts were less frequent and not always in the favorable direction. At the end of the age 4 year (the year before kindergarten entry), two statistically significant impacts were found for children in the 3-year-old cohort: children in the Head Start group scored higher than children in the non-Head Start group on the CTOPPP Elision measure as well as on parental reports of their child's literacy skills. However, in kindergarten, teachers reported poorer math skills for children in the Head Start group than for those in the control group. At the end of 1st grade, there was suggestive evidence of a favorable impact on oral comprehension. Yet, at the end of 3rd grade, there was only one significant impact, and it was unfavorable—the parents of the Head Start group children reported a significantly lower child promotion rate than the parents of the non-Head Start group children.

Exhibits 4.12 and 4.13 summarize the statistically significant cognitive impacts and their effect sizes through the end of 3rd grade.

Impacts on Social-Emotional Outcomes

In the social-emotional domain, the results differed by age cohort and by the source of the information on the child's behavior. For children in the 4-year-old cohort, there were no observed impacts at the end of Head Start or at the end of kindergarten. In elementary school, statistically significant impacts emerged, but there was a marked difference in how parents' and teachers' reports compared for the treatment and control groups. At the end of 1st grade, there was a favorable impact on parent reports of children's withdrawn behavior (a lower incidence for Head Start group children), but an unfavorable impact on teachers' reports of the children's behavior in class, i.e., a greater incidence of reticent behavior and problematic teacher-child interactions for Head Start group children. A similar pattern was observed at the end of 3rd grade, with favorable impacts on parents' reports of children's aggressive behavior and total behavior problems, but unfavorable impacts on teachers' reports of children's emotional symptoms, closeness with teacher, and teacher-child relationships. In addition, there was an unfavorable impact on children's self-reports of their relationships with school peers. The study collected children's self-reports for the first time at the end of 3rd grade.

Exhibit 4.12. Summary of ITT Cognitive Impacts for 4-Year-Olds by Year

Measure	Age 4 (Head Start Year)	K	1st Grade	3rd Grade
Language, Literacy, and Pre-Writing				
Color Identification	**0.16**	NA	NA	NA
Pre-Writing (McCarthy Draw a Design)		NA	NA	NA
Emergent Literacy Scale (parent report)	**0.31**	NA	NA	NA
Letter Naming	**0.25**		NA	NA
Test of Phonological Processing (CTOPPP Elision)			NA	NA
Receptive Vocabulary (PPVT)	**0.09**		**0.09**	
Letter-Word Identification (WJIII)	**0.22**			
Spelling (WJIII)	**0.15**			NA
Oral Comprehension (WJIII)				NA
Pre-Academic Skills (WJIII)	**0.19**			NA
Phonetic Skills/ Word Attack (WJIII)	NA			NA
Basic Reading (WJIII)	NA			NA
Academic Applications (WJIII)	NA	NA		NA
Academic Skills (WJIII)	NA	NA		NA
Passage Comprehension (WJIII)	NA	NA		NA
ECLS-K Reading	NA	NA		**0.11**
Writing Sample (WJIII)	NA	NA		NA
Spanish Language				
Receptive Vocabulary (TVIP)				NA
Batería WM Identificación de letras y palabras				
Math				
One-to-One Counting (Counting Bears)		NA	NA	NA
Applied Problems (WJIII)				
Quantitative Concepts (WJIII)	NA			NA
Math Reasoning (WJIII)	NA			NA
Calculation (WJIII)	NA	NA		
School Performance				
School Accomplishments	NA			NA
Promotion (parent report)	NA			
Language and Literacy Ability	NA			
Math Ability	NA			
Math Skills	NA	NA	NA	
Reading/Language Arts Skills	NA	NA	NA	
Social Studies and Science Ability	NA			NA

KEY:

Dark gray cell indicates a significant favorable impact ($p \leq 0.10$).

Blank cell indicates no significant impact.

NA indicates that no data was collected for this outcome at this data collection point.

NOTE: Intent to Treat (ITT) effect sizes are presented only for statistically significant differences ($p \leq 0.10$). The effect size is simply the impact estimate divided by the standard deviation of the outcome measure in the population. The effect size provides an indication of the magnitude of each impact that is independent of the particular instrument or measure used. More discussion of the interpretation of effect sizes is provided in Chapter 2.

Exhibit 4.13. Summary of ITT Cognitive Impacts for 3-Year-Olds by Year

Measure	Age 3 (Head Start Year)	Age 4	K	1st Grade	3rd Grade
Language, Literacy, and Pre-Writing					
Color Identification			NA	NA	NA
Pre-Writing (McCarthy Draw a Design)	**0.14**		NA	NA	NA
Emergent Literacy Scale (parent report)	**0.35**	**0.16**	NA	NA	NA
Letter Naming	**0.24**			NA	NA
Test of Phonological Processing (CTOPPP Elision)	**0.10**	**0.15**		NA	NA
Receptive Vocabulary (PPVT)	**0.18**				
Letter-Word Identification (WJIII)	**0.26**				
Spelling (WJIII)					NA
Oral Comprehension (WJIII)				**0.08**	NA
Pre-Academic Skills (WJIII)	**0.22**				NA
Phonetic Skills/Word Attack (WJIII)	NA	NA			NA
Basic Reading (WJIII)	NA	NA			NA
Academic Applications (WJIII)	NA	NA	NA		NA
Academic Skills (WJIII)	NA	NA	NA		NA
Passage Comprehension (WJIII)	NA	NA	NA		NA
ECLS-K Reading	NA	NA	NA		
Writing Sample (WJIII)	NA	NA	NA		NA
Spanish Language					
Receptive Vocabulary (TVIP)					NA
Batería WM Identificación de letras y palabras			**0.26**		
Math					
One-to-One Counting/Counting Bears			NA	NA	NA
Applied Problems (WJIII)	**0.15**				
Quantitative Concepts (WJIII)	NA	NA			NA
Math Reasoning (WJIII)	NA	NA			NA
Calculation (WJIII)	NA	NA	NA		
School Performance					
School Accomplishments	NA	NA			NA
Promotion (parent report)	NA	NA			**-0.11**
Language and Literacy Ability	NA	NA			
Math Ability	NA	NA	**-0.19**		
Math Skills	NA	NA	NA	NA	
Reading/Language Arts Skills	NA	NA	NA	NA	
Social Studies and Science Ability	NA	NA			NA

KEY:

Dark gray cell indicates a significant favorable impact ($p \leq 0.10$).

Diagonal mark cell indicates a significant unfavorable impact ($p \leq 0.10$).

Blank cell indicates no significant impact.

NA indicates that no data were collected for this outcome at this data collection point.

NOTE: Intent to Treat (ITT) effect sizes are presented only for statistically significant differences ($p \leq 0.10$). The effect size is simply the impact estimate divided by the standard deviation of the outcome measure in the population. The effect size provides an indication of the magnitude of each impact that is independent of the particular instrument or measure used. More discussion of the interpretation of effect sizes is provided in Chapter 2.

In contrast, for children in the 3-year-old cohort, there were favorable impacts at all data collection points through the end of 3rd grade on several parent-reported measures of children's social-emotional development. There were no impacts on teacher-reported measures of social-emotional development for the 3-year-old cohort at any data collection point or on the children's self-reports in 3rd grade. During the Head Start year, children in the 3-year-old cohort experienced favorable impacts on overall problem behavior and hyperactivity and on their social skills and positive approaches to learning at the end of their age 4 year. The favorable impacts on social skills and approaches to learning, and on hyperactivity, were also observed at the end of kindergarten. At the end of 1st grade, there were favorable impacts on parents' reports of positive relationship and closeness in the parent-child relationship. Finally, at the end of 3rd grade, there was one favorable impact on social skills and approaches to learning.

Exhibits 4.14 and 4.15 summarize the statistically significant social-emotional impacts and their effect sizes through the end of 3rd grade.

Impacts on Health Outcomes

On measures in the health domain, there were large effects noted during the preschool years for both age cohorts on the receipt of dental care and, for children in the 3-year-old cohort, on parental reports of their child's overall health status. Positive impacts on access to health insurance coverage were found for children in the 4-year-old cohort at the end of kindergarten and 1st grade and at the end of kindergarten for children in the 3-year-old cohort. However, by the end of the 3rd grade, there were no remaining impacts on any of the health outcomes for either age cohort.

Exhibits 4.16 and 4.17 summarize the statistically significant health impacts and their effect sizes through the end of 3rd grade.

Exhibit 4.14. Summary of ITT Social-Emotional Impacts for 4-Year-Olds by Year

Measure	Age 4 (Head Start Year)	K	1st Grade	3rd Grade
Parent-Reported Measures				
Aggressive Behavior				**-0.13**
Hyperactive Behavior				
Withdrawn Behavior			**-0.13**	
Total Problem Behavior				**-0.12**
Social Competencies				NA
Social Skills and Positive Approaches To Learning				
Closeness with Parent				NA
Conflict with Parent				NA
Positive Parent-Child Relationships				NA
Teacher-Reported Measures				
Aggressive (ASPI)	NA			NA
Interactive/Hyperactive (ASPI)	NA			
Withdrawn/Low Energy (ASPI)	NA			NA
Oppositional (ASPI)	NA			NA
Problems with Peer Interaction (ASPI)	NA			NA
Shy/Socially Reticent (ASPI)	NA		**0.19**	NA
Problems with Structured Learning (ASPI)	NA			NA
Problems with Teacher Interaction (ASPI)	NA		**0.13**	NA
Closeness with Teacher	NA			**-0.13**
Conflict with Teacher	NA			
Positive Teacher-Child Relationships	NA			**-0.14**
Conduct Problems (% in Normal Category)	NA	NA	NA	
Emotional Symptoms (% in Normal Category)	NA	NA	NA	**-0.24**
Hyperactivity (% in Normal Category)	NA	NA	NA	
Peer Problems (% in Normal Category)	NA	NA	NA	
Pro-social Behavior (% in Normal Category)	NA	NA	NA	
Total Difficulties (% in Normal Category)	NA	NA	NA	
Social Competency	NA	NA	NA	
Child-Reported Measures				
Externalizing	NA	NA	NA	
Internalizing	NA	NA	NA	
Peer Relations	NA	NA	NA	**-0.13**
School	NA	NA	NA	

KEY:

Dark gray cell indicates a significant favorable impact ($p \leq 0.10$).

Diagonal mark cell indicates a significant unfavorable impact ($p \leq 0.10$).

Blank cell indicates no significant impact.

NA indicates that no data were collected for this outcome at this data collection point.

NOTE: Intent to Treat (ITT) effect sizes are presented only for statistically significant differences ($p \leq 0.10$). The effect size is simply the impact estimate divided by the standard deviation of the outcome measure in the population. The effect size provides an indication of the magnitude of each impact that is independent of the particular instrument or measure used. More discussion of the interpretation of effect sizes is provided in Chapter 2.

Exhibit 4.15. Summary of ITT Social-Emotional Impacts for 3-Year-Olds by Year

Measure	Age 3 (Head Start Year)	Age 4	K	1st Grade	3rd Grade
Parent-Reported Measures					
Aggressive Behavior					
Hyperactive Behavior	**-0.21**		**-0.12**		
Withdrawn Behavior					
Total Problem Behavior	**-0.14**				
Social Competencies					NA
Social Skills and Positive Approaches To Learning		**0.11**	**0.14**		**0.12**
Closeness with Parent				**0.10**	NA
Conflict with Parent					NA
Positive Parent-Child Relationships				**0.10**	NA
Teacher-Reported Measures					
Aggressive (ASPI)	NA	NA			NA
Interactive/Hyperactive (ASPI)	NA	NA			
Withdrawn/Low Energy (ASPI)	NA	NA			NA
Oppositional (ASPI)	NA	NA			NA
Problems with Peer Interaction (ASPI)	NA	NA			NA
Shy/Socially Reticent (ASPI)	NA	NA			NA
Problems with Structured Learning (ASPI)	NA	NA			NA
Problems with Teacher Interaction (ASPI)	NA	NA			NA
Closeness with Teacher	NA	NA			
Conflict with Teacher	NA	NA			
Positive Teacher-Child Relationships	NA	NA			
Conduct Problems (% in Normal Category)	NA	NA	NA	NA	
Emotional Symptoms (% in Normal Category)	NA	NA	NA	NA	
Hyperactivity (% in Normal Category)	NA	NA	NA	NA	
Peer Problems (% in Normal Category)	NA	NA	NA	NA	
Pro-social Behavior (% in Normal Category)	NA	NA	NA	NA	
Total Difficulties (% in Normal Category)	NA	NA	NA	NA	
Social Competency	NA	NA	NA	NA	
Child-Reported Measures					
Externalizing	NA	NA	NA	NA	
Internalizing	NA	NA	NA	NA	
Peer Relations	NA	NA	NA	NA	
School	NA	NA	NA	NA	

KEY:

Dark gray cell indicates a significant favorable impact ($p \leq 0.10$).

Blank cell indicates no significant impact.

NA indicates that no data was collected for this outcome at this data collection point.

NOTE: Intent to Treat (ITT) effect sizes are presented only for statistically significant differences ($p \leq 0.10$). The effect size is simply the impact estimate divided by the standard deviation of the outcome measure in the population. The effect size provides an indication of the magnitude of each impact that is independent of the particular instrument or measure used. More discussion of the interpretation of effect sizes is provided in Chapter 2.

Exhibit 4.16. Summary of ITT Health Impacts for 4-Year-Olds by Year

Measure	Age 4 (Head Start Year)	K	1st Grade	3rd Grade
Parent-Reported Measures				
Child Received Dental Care	**0.31**			
Child Has Health Insurance Coverage		**0.11**	**0.11**	
Child's Overall Health Status Is Excellent/Good		**0.13**		
Child Needs Ongoing Health Care				
Child Had Care for Injury in Last Month				

KEY:

Dark gray cell indicates a significant favorable impact ($p \leq 0.10$).

Blank cell indicates no significant impact.

NA indicates that no data was collected for this outcome at this data collection point.

NOTE: Intent to Treat (ITT) effect sizes are presented only for statistically significant differences ($p \leq 0.10$). The effect size is simply the impact estimate divided by the standard deviation of the outcome measure in the population. The effect size provides an indication of the magnitude of each impact that is independent of the particular instrument or measure used. More discussion of the interpretation of effect sizes is provided in Chapter 2.

Exhibit 4.17. Summary of ITT Health Impacts for 3-Year-Olds by Year

Measure	Age 3 (Head Start Year)	Age 4	K	1st Grade	3rd Grade
Parent-Reported Measures					
Child Received Dental Care	**0.33**	**0.20**			
Child Has Health Insurance Coverage			**0.14**		
Child's Overall Health Status Is Excellent/Good	**0.11**				
Child Needs Ongoing Health Care					
Child Had Care for Injury in Last Month[65]		**0.10***			

KEY:

Dark gray cell indicates a significant favorable impact ($p \leq 0.10$).

Blank cell indicates no significant impact.

An * indicates that the interpretation of the outcome is unclear.

NA indicates that no data was collected for this outcome at this data collection point.

NOTE: Intent to Treat (ITT) effect sizes are presented only for statistically significant differences ($p \leq 0.10$). The effect size is simply the impact estimate divided by the standard deviation of the outcome measure in the population. The effect size provides an indication of the magnitude of each impact that is independent of the particular instrument or measure used. More discussion of the interpretation of effect sizes is provided in Chapter 2.

[65] The interpretation of child had care for injury in last month is unclear. The change may reflect an increase in injuries, an increase in care seeking, or both.

Impacts on Parenting Practices

In the area of parenting practices, the results differed by age cohort. For children in the 4-year-old cohort, there was a reduction on parents' use of time out[66] for discipline at the end of the Head Start year, and no significant impacts at the end of kindergarten or at the end of 1st grade. However, by the end of the 3rd grade, there was strong evidence of a large positive impact on parental reports of the amount of time they spent with their child.

For children in the 3-year-old cohort, there were favorable impacts on multiple outcomes during preschool, including less spanking and less use of an authoritarian parenting style (characterized by low warmth and high parental control), and increases in reading to their child and participation in cultural enrichment activities. There was a reduction in spanking and time out at the end of kindergarten, as well as a reduction in the use of time out and the authoritarian parenting style at the end of 1st grade. At the end of the 3rd grade there was a positive impact on the use of the preferred authoritative parenting style (characterized by high warmth and high control).

Exhibits 4.18 and 4.19 summarize the statistically significant parenting practices impacts and their effect sizes through the end of 3rd grade.

[66] The interpretation of time out is unclear. This change may reflect favorable changes in the children's behavior, changes in the parents' reactions (whether to less or more desirable forms of discipline), or both.

Exhibit 4.18. Summary of ITT Parenting Practices Impacts for 4-Year-Olds by Year

Measure	Age 4 (Head Start Year)	K	1st Grade	3rd Grade
Parent-Reported Measures				
Parent Spanked Child in Last Week				
Parent Used Time Out in Last Week[67]	**-0.17***			NA
Parent Read to Child in Last Week				NA
Parental Safety Practices Scale				NA
Family Cultural Enrichment Scale				NA
Parenting Style: Authoritarian	NA			
Parenting Style: Authoritative	NA			
Parenting Style: Neglectful	NA			
Parenting Style: Permissive	NA			
Supportive School Environment	NA	NA	NA	
Effect of Parenting on Parent's Life	NA	NA	NA	
Doing Things Together	NA	NA	NA	
Time Spent with Child	NA	NA	NA	**0.27**
Parent Perception of School Services	NA	NA	NA	
Teacher-Reported Measures				
School Contact and Communication	NA			
Parent Participation	NA			

KEY:

Dark gray cell indicates a significant favorable impact ($p \leq 0.10$).

Blank cell indicates no significant impact.

An * indicates that the interpretation of the outcome is unclear.

NA indicates that no data was collected for this outcome at this data collection point.

NOTE: Intent to Treat (ITT) effect sizes are presented only for statistically significant differences ($p \leq 0.10$). The effect size is simply the impact estimate divided by the standard deviation of the outcome measure in the population. The effect size provides an indication of the magnitude of each impact that is independent of the particular instrument or measure used. More discussion of the interpretation of effect sizes is provided in Chapter 2.

[67] See footnote 65.

Exhibit 4.19. Summary of ITT Parenting Practices Impacts for 3-Year-Olds by Year

Measure	Age 3 (Head Start Year)	Age 4	K	1st Grade	3rd Grade
Parent-Reported Measures					
Parent Spanked Child in Last Week	**-0.14**		**-0.09**		
Parent Used Time Out in Last Week[68]			**-0.13***	**-0.11***	NA
Parent Read to Child in Last Week	**0.15**				NA
Parental Safety Practices Scale					NA
Family Cultural Enrichment Scale	**0.18**				NA
Parenting Style: Authoritarian	NA	**-0.14**		**-0.11**	
Parenting Style: Authoritative	NA				**0.16**
Parenting Style: Neglectful	NA				
Parenting Style: Permissive	NA				
Supportive School Environment	NA	NA	NA	NA	
Effect of Parenting on Parent's Life	NA	NA	NA	NA	
Doing Things Together	NA	NA	NA	NA	
Time Spent with Child	NA	NA	NA	NA	
Parent Perception of School Services	NA	NA	NA	NA	
Teacher-Reported Measures					
School Contact and Communication	NA	NA			
Parent Participation	NA	NA			

KEY:

Dark gray cell indicates a significant favorable impact ($p \leq 0.10$).

Blank cell indicates no significant impact.

An * indicates that the interpretation of the outcome is unclear.

NA indicates that no data was collected for this outcome at this data collection point.

NOTE: Intent to Treat (ITT) effect sizes are presented only for statistically significant differences ($p \leq 0.10$). The effect size is simply the impact estimate divided by the standard deviation of the outcome measure in the population. The effect size provides an indication of the magnitude of each impact that is independent of the particular instrument or measure used. More discussion of the interpretation of effect sizes is provided in Chapter 2.

Summary

In the cognitive domain, the Head Start Impact Study showed impacts for the sample as a whole at the end of one year of Head Start on a broad range of early language and literacy outcomes for children in both the 3- and 4-year-old cohorts. However, these early effects rapidly dissipated in elementary school, with only a single impact remaining at the end of 3rd grade for children in each age cohort: a favorable impact for the 4-year-old cohort and an unfavorable impact for the 3-year-old cohort. In the social-emotional domain, the results differed by age

[68] See footnote 65.

cohort and by the source of the information on the child's behavior. For children in the 4-year-old cohort, there were no observed impacts through the end of kindergarten and then favorable impacts reported by parents and unfavorable impacts reported by teachers at the end of 1st and 3rd grades. Children's own reports showed one unfavorable impact at the end of 3rd grade. In contrast, for children in the 3-year-old cohort, there were favorable impacts at all data collection points through the end of 3rd grade on several parent-reported measures of children's social-emotional development. However, there were no impacts on teacher-reported measures of social-emotional development for the 3-year-old cohort at any data collection point or on the children's self-reports in 3rd grade.

In the health domain, early favorable impacts were noted for both age cohorts, but by the end of 3rd grade, there we no remaining impacts for either age cohort. For the 4-year-old cohort in the parenting practices domain, there were few favorable impacts, with one impact reported at the end of the Head Start year and one impact at the end of the 3rd grade. In contrast, for the 3-year-old cohort, there were favorable parent-reported impacts across all years of the study.

Chapter 5: Variation in Impacts: Child and Parent Subgroups

Introduction

Previous chapters presented estimates for the effect of the national Head Start system on children who entered the program in fall 2002 in non-saturated communities (i.e., communities whose Head Start programs were not undersubscribed[69]). This chapter looks instead for differences in impacts among different types of children and parents to respond to the congressional mandate that the Head Start Impact Study look for *". . .possible sources of variation in impact of the Head Start program."*

For this analysis, seven dimensions were used to define subgroups: (1) child's level of pre-academic skills at the start of Head Start (children in the lowest quartile vs. other children), (2) child's status as a Dual Language Learner (DLL) at the start of Head Start, (3) child's special needs (as reported by the parent at the start of Head Start), (4) biological mother's/caregiver's race/ethnicity, (5) reported level of depressive symptoms for the child's parent/caregiver, (6) a composite index of household risks, and (7) urbanicity. These dimensions were based on data collected at the time of random assignment.

This chapter concentrates on results that meet two criteria: (1) a statistically significant difference in impacts, i.e., where there was a statistically significant difference in Head Start's effects for one subgroup compared to another (e.g., impacts were found to be statistically higher for children in the lowest quartile on pre-academic skills than for children not in the lowest quartile) and (2) a statistically significant impact on one of the individual subgroups (e.g., a statistically significant impact on children in the lowest quartile, a significant impact on children not in the lowest quartile, or significant impacts on both groups). For this determination we used a standard of statistical significance of $p \leq 0.10$. The results of pre-K through 3rd grade analyses are provided at the end of the chapter (these tables also note which effects passed the Benjamini-Hochberg test for multiple comparisons). All the subgroup findings, including effect sizes, are available on the Administration for Children and Families, Office of Planning, Research and Evaluation Website at: http://www.acf.hhs.gov/.

[69] These communities served 85 percent of the newly-entering children in the nation that year.

Subgroups Examined

With regard to participant characteristics, the following seven dimensions were identified as being of primary policy interest and used to create subgroups (all were measured at baseline):

- **Child's Pre-Academic Skills—**based on whether the child scored in the lowest quartile of the study population on the baseline assessment of the Woodcock-Johnson III Pre-Academic Skills (comprising three tests: Letter-Word Identification, Spelling, and Applied Problems). Two subgroups were created using this test score: the child was in the lowest quartile subgroup, or the child was not in the lowest quartile subgroup.

- **Child's Home Language—**based on the language in which the child was assessed for the baseline assessment in fall 2002. Two subgroups were created: the child was English speaking, or the child was a Dual Language Learner (See Chapter 2 in the *Head Start Impact Study Final Report*[70] for how the language for the baseline assessment was determined.)

- **Special Needs—**based on the parent's response to the following question on the baseline interview, "Did a doctor or other health or education professional ever tell you that [CHILD] has any special needs or disabilities—for example, physical, emotional, language, hearing, learning difficulty, or other special needs?" Two subgroups were created: the child was reported to have special needs, or the child was not reported to have special needs.

- **Biological Mother/Caregiver Race/Ethnicity—**based on the race and ethnicity of the person identified as being most responsible for the care of the child at the time of the baseline parent interview.[71] Three categories were created: non-Hispanic White or other,[72] non-Hispanic Black, and Hispanic.

- **Parent/Caregiver-Reported Depressive Symptoms—**determined from responses to the baseline parent/caregiver interview using the shortened version (12 items) of the Center for Epidemiologic Studies-Depression scale (CES-D) (Seligman, 1993[73]). Four subgroups were created from the scale: (1) no depressive symptoms (score of 0-4), (2) mild depressive symptoms (score of 5-9), (3) moderate depressive symptoms (score of 10-14), and (4) severe depressive symptoms (score of 15-36).

---

[70] http://www.acf.hhs.gov/programs/opre/hs/impact_study/reports/impact_study_final.pdf

[71] The primary caregiver is the child's biological mother for 96 percent of the study children.

[72] Other race (N=94 for the 3-year-old cohort and N=85 for the 4-year-old cohort) was combined with White because the number of other race respondents was too small to study independently.

[73] The four depressive symptoms categories are reported on page 101 in the above reference for the 20 item CES-D. The cut points were proportionately adjusted for the shortened version of the CES-D for use in ECLS-B, FACES, and HSIS.

- **Household risk index**—determined by the number of the following characteristics reported in the baseline parent interview: (1) receipt of TANF or Food Stamps, (2) neither parent in household has high school diploma or a GED, (3) neither parent in household is employed or in school, (4) the child's biological mother/caregiver is a single parent, and (5) the child's biological mother was age 18 or younger when child was born. A child's family score could range from 0 to 5 points. Three categories were created: low/no risk (0-2 risk factors), moderate risk (3 risk factors), and high risk (4-5 risk factors).

- **Urbanicity**—based on the location of the Head Start center at which the family applied for admission. If the center was located in a Census-defined urbanized area, the family was considered to live in an urban area; if not, the family was considered not to live in an urban area. Thus, two subgroups were defined.

Exhibit 5.1 provides the distribution of children across the subgroups by age cohort and status as a part of the Head Start group or the control group. There were no significant differences between the two groups.

Exhibit 5.1. Distribution of Children Across the Subgroups by Age Cohort and Random Assignment Status

Subgroup	4-Year-Old Cohort		3-Year-Old Cohort	
	Head Start Group	Control Group	Head Start Group	Control Group
Child's Pre-Academic Skills				
Lowest Quartile	249 (24%)	159 (26%)	312 (24%)	189 (25%)
Non-lowest Quartile	775 (76%)	455 (74%)	987 (76%)	582 (75%)
Child's Home Language				
English Speaking	695 (68%)	418 (68%)	996 (77%)	593 (77%)
Dual Language Learners	329 (32%)	196 (32%)	303 (23%)	178 (23%)
Special Needs				
Special Needs	146 (14%)	74 (12%)	171 (13%)	77 (10%)
Non-special Needs	878 (86%)	540 (88%)	1,128 (87%)	694 (90%)
Biological Mother/Caregiver Race/Ethnicity				
Non-Hispanic White Children/Other	374 (37%)	220 (36%)	396 (31%)	276 (36%)
Non-Hispanic Black Children	229 (22%)	134 (22%)	486 (37%)	256 (33%)
Hispanic Children	421 (41%)	260 (42%)	417 (32%)	239 (31%)
Parent/Caregiver Reported Depressive Symptoms				
No Depressive Symptoms	478 (50%)	279 (51%)	635 (53%)	377 (56%)
Mild Depressive Symptoms	250 (26%)	144 (27%)	310 (26%)	155 (23%)
Moderate Depressive Symptoms	127 (13%)	65 12%)	149 (12%)	83 (12%)
Severe Depressive Symptoms	98 (10%)	55 (10%)	107 (9%)	62 (9%)
Household Risk Index				
Low/No Household Risk	744 (73%)	456 (74%)	956 (74%)	568 (74%)
Moderate Household Risk	204 (20%)	110 (18%)	234 (18%)	145 (19%)
High Household Risk	76 (7%)	48 (8%)	109 (8%)	58 (7%)

Exhibit 5.1. Distribution of Children Across the Subgroups by Age Cohort and Random Assignment Status (continued)

Subgroup	4-Year-Old Cohort		3-Year-Old Cohort	
	Head Start Group	Control Group	Head Start Group	Control Group
Urbanicity				
Urban	872 (85%)	530 (86%)	1,077 (83%)	629 (82%)
Not Urban	152 (15%)	84 (14%)	222 (17%)	142 (18%)

Note: Numbers are based on the spring 2003 analysis sample. Due to rounding, the sum of the percents may not equal to 100 percent.

Analytic Approach

The determination of what constitutes a pattern of differential impacts between subgroups is not simple, and there is no scientific consensus on how best to make these determinations. Each of the seven dimensions was comprised of at least two separate subgroups of children (e.g., children with and without special needs), and there are at least three statistical tests conducted on each outcome for each subgroup: (e.g., *within subgroup* impacts in this example measured impacts for children with and without special needs separately, as well as a test of the difference in impacts between children with and without special needs, the *difference of difference* test[74]). For dimensions that have more than two subgroups of children, such as race/ethnicity, the number of tests was even greater. All subgroups were analyzed for the 3- and 4-year-old cohorts separately.

Given the many outcome measures, the multiple years of data collection, the two age cohorts, and the seven subgroup dimensions, 13,668 statistical tests involving subgroups were conducted. When so many statistical tests and analyses are run, it is important to guard against Type I errors, statistically significant findings that reflect chance variations rather than true differences. As discussed in Chapter 2, the Benjamini-Hochberg procedure was used to address this problem. Findings from the Benjamin-Hochberg procedure, limiting the discovery rate to at most 10 percent, are highlighted in the exhibits. However, since subgroups are smaller in size and therefore have lower power than the full sample to detect true effects, the risk of Type II

[74] This is referred to as a difference of difference test because of the nature of the comparison. For each subgroup within the given subgroup dimension—children with and without special needs—there is first a test of the *within-group impact.* This test analyzes the difference between the Head Start and control groups for special needs children only or for non-special needs children only. Then these within-group impacts are compared to one another, resulting in a test of the difference of those impacts. The difference in difference test tells us whether an impact is significantly larger for one subgroup than for another.

error (i.e., failing to detect true subgroup impacts or true differences in subgroup impacts that do occur) is greater, and the use of multiple comparison procedures increases the risk of a Type II error. Due to these limitations, the subgroup findings should be viewed as secondary and exploratory as compared to the main impact findings that are considered primary as well as confirmatory.

We have aimed to make this chapter useful and readable by being selective in determining the results to highlight in the narrative. Accordingly, the discussion primarily focuses on results where there was both a statistically significant difference in impacts between subgroups and a statistically significant impact for at least one subgroup in the comparison. End of 3rd grade results are presented by cohort for each subgroup dimension. At the end of the chapter, we discuss how subgroup findings in 3rd grade relate to prior time points.

End of 3rd Grade Results

Child's Pre-Academic Skills

Findings for subgroups defined by pre-academic skills (i.e., children in the lowest academic quartile/those not in the lowest academic quartile) are presented in Exhibits 5.2 and 5.3. For the 4-year-old cohort, there is no clear pattern, despite several statistically significant differences in impacts across these two groups (Exhibit 5.2.) Where there were statistically significant differences in impacts, the direction was mixed within each of the groups. In other words, there were some favorable and some unfavorable impacts for children in the lowest academic quartile and for children who were not in the lowest academic quartile. Children in the lowest quartile showed a mix of favorable and unfavorable impacts on several parent-reported measures of parenting and perceptions of the elementary school, including reductions in the use of neglectful parenting styles, but also reductions in the perceived supportiveness of the school environment and perceptions of school services. For children who were not in the lowest quartile at baseline, children in the Head Start group reported poorer peer relations than children in the control group, while parents in the Head Start group reported a more supportive school environment.

For the 3-year-old cohort, there was only one differential impact. There was an unfavorable impact on parents' reports of their children's withdrawn behavior for children who

were not in the lowest quartile at baseline. In contrast, there was no impact on the withdrawn behavior of children who were in the lowest quartile at baseline.

Exhibit 5.2. Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Child's Pre-Academic Skills

Outcome Measure (Number of Outcomes Examined)	Impact in Subgroup A (Head Start – Control) Children Not in Lowest Quartile	Impact in Subgroup B (Head Start – Control) Children in Lowest Quartile	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE			
Direct Child Assessment (5)			
School Performance[75] (5)			
SOCIAL-EMOTIONAL			
Parent Report (5)			
Teacher Report (10)			
Child Report (4)			
Peer Relations	**-0.21*** **	0.08	A-B*
HEALTH (5)			
PARENTING			
Parent Report (10)			
Supportive School Environment	0.14*	**-0.30** **	**A-B*** **
Parenting Style: Neglectful	0.03	**-0.12** **	**A-B*** **
Parent Perception of School Services	-0.01	**-0.31** **	A-B*
Teacher Report (2)			

Key:
*******$p \leq 0.01$
******$p \leq 0.05$
*****$p \leq 0.10$

Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.

[75] For all tables in this chapter, school performance measures are teacher-reported except for promotion which is parent-reported.

Exhibit 5.3. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Child's Pre-Academic Skills

Outcome Measure (Number of Outcomes Examined)	Impact in Subgroup A (Head Start – Control) Children Not in Lowest Quartile	Impact in Subgroup B (Head Start – Control) Children in Lowest Quartile	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE			
Direct Child Assessment (5)			
School Performance (5)			
SOCIAL-EMOTIONAL			
Parent Report (5)			
Withdrawn Behavior	0.20**	-0.24	A-B**
Teacher Report (10)			
Child Report (4)			
HEALTH (5)			
PARENTING			
Parent Report (10)			
Teacher Report (2)			

Key:
***$p \leq 0.01$
**$p \leq 0.05$
*$p \leq 0.10$

Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.

Children with Special Needs

As shown in Exhibits 5.4 and 5.5, there were few statistically significant differences in impacts for children with and without special needs. For the 4-year-old cohort, there was one statistically significant differential impact in letter-word identification skills, with a favorable impact for those who did not have a special need at baseline and no statistically significant impact for those who did. For the 3-year old cohort, no significant findings were found for the subgroups defined by special needs.

Exhibit 5.4. Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Child's Special Needs

Outcome Measure (Number of Outcomes Examined)	Impact in Subgroup A (Head Start – Control) Children Without Special Needs	Impact in Subgroup B (Head Start – Control) Children With Special Needs	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE			
Direct Child Assessment (5)			
WJ III Letter-Word Identification	3.71*	-8.66	A-B*
School Performance (5)			
SOCIAL-EMOTIONAL			
Parent Report (5)			
Teacher Report (10)			
Child Report (4)			
HEALTH (5)			
PARENTING			
Parent Report (10)			
Teacher Report (2)			

Key:
***$p \leq 0.01$
**$p \leq 0.05$
*$p \leq 0.10$

Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.

Exhibit 5.5. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Child's Special Needs

Outcome Measure (Number of Outcomes Examined)	Impact in Subgroup A (Head Start – Control) Children Without Special Needs	Impact in Subgroup B (Head Start – Control) Children With Special Needs	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE			
Direct Child Assessment (5)			
School Performance (5)			
SOCIAL-EMOTIONAL			
Parent Report (5)			
Teacher Report (10)			
Child Report (4)			
HEALTH (5)			
PARENTING			
Parent Report (10)			
Teacher Report (2)			

Key:
*******$p \leq 0.01$
******$p \leq 0.05$
*****$p \leq 0.10$

Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.

Dual Language Learners

Exhibits 5.6 and 5.7 show the differences in the effect of Head Start for subgroups defined by primary home language for the 4-year-olds and 3-year-olds, respectively. There were few statistically significant differential impacts for children in either age cohort based on subgroups defined by home language. For the 4-year-old cohort, there was a favorable impact for Dual Language Learners with respect to parent perceptions of the supportiveness of the school environment. Head Start parents of Dual Language Learners perceived the school environment as more supportive than control group parents of Dual Language Learners. There was no

statistically significant impact on this variable for parents of children who were not Dual Language Learners.

Exhibit 5.6. Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Child's Home Language

Outcome Measure (Number of Outcomes Examined)	Impact in Subgroup A (Head Start – Control) Dual Language Learners	Impact in Subgroup B (Head Start – Control) English-Speaking Children	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE			
Direct Child Assessment (5)			
School Performance (5)			
SOCIAL-EMOTIONAL			
Parent Report (5)			
Teacher Report (10)			
Child Report (4)			
HEALTH (5)			
PARENTING			
Parent Report (10)			
Supportive School Environment	**0.35*** **	-0.11	**A-B*** **
Teacher Report (2)			

Key:
*******$p \leq 0.01$
******$p \leq 0.05$
*****$p \leq 0.10$

Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.

Exhibit 5.7. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Child's Home Language

Outcome Measure (Number of Outcomes Examined)	Impact in Subgroup A (Head Start – Control) Dual Language Learners	Impact in Subgroup B (Head Start – Control) English-Speaking Children	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE			
Direct Child Assessment (5)			
School Performance (5)			
SOCIAL-EMOTIONAL			
Parent Report (5)			
Social Skills and Positive Approaches to Learning	-0.06	0.31***	A-B*
Teacher Report (10)			
Child Report (4)			
HEALTH (5)			
Child's Overall Health is Excellent/Good	**0.17*** **	**-0.06** **	**A-B*** **
PARENTING			
Parent Report (10)			
Teacher Report (2)			

Key:
***p ≤ 0.01
**p ≤ 0.05
*p ≤ 0.10

Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.

There were two differential impacts for the 3-year-old cohort, and they were mixed in direction. DLLs showed a favorable impact of Head Start on parents' reports of children's health, whereas there was an unfavorable impact on this outcome for English-speaking children At the same time, however, parents of English speaking children reported more favorable assessments of their children's social skills and positive approaches to learning than their control group peers, whereas there was no statistically significant impact on this variable for children who were Dual Language Learners.

Race/Ethnicity

Findings for subgroups defined by biological mother/caregiver race/ethnicity subgroup are presented in Exhibits 5.8 (4-year-old cohort) and 5.9 (3-year-old cohort). Favorable impacts at the end of 3rd grade were found for Black children in the 4-year-old cohort in the social-emotional and parenting domains. According to parents, Black children in the Head Start group had less aggressive and hyperactive behavior and fewer total behavior problems than their peers in the control group. In addition, Black children in the Head Start group reported more favorable school experiences than their control group peers. In the parenting domain, Black parents in the Head Start group reported they were less likely to spank their children than Black parents in the control group.

In contrast to the Black group, the Hispanic group showed mixed results, and the White group showed several unfavorable impacts. Hispanic Head Start group children's self-reports showed unfavorable impacts on internalizing behavior (more likely to be sad, lonely or anxious).[76] The Head Start Hispanic parents viewed the school environment as more supportive than control group parents.

Among White children, teachers reported less closeness with children in the Head Start group than children in the control group. They also reported that parents of children in the Head Start group were less likely to participate in school activities than children in the control group among White families. White children themselves reported that they had poorer relationships with their peers if they were in the Head Start group than in the control group.

Results for the 3-year-old cohort showed mixed results across the different racial and ethnic groups. Where differential impacts existed, White children showed a mix of favorable and unfavorable impacts. White children in the Head Start group were less likely to be promoted to the next grade, and their parents reported more aggressive behavior than their control group counterparts. They were also, however, reported to have more social skills and positive approaches to learning according to their parents, and their parents were more likely to participate in school according to teachers. Additionally, White Head Start parents reported they were more likely to spank their children than White control group parents.

---

[76] For the directionality of social-emotional measures, see Chapter 4 in this report.

Exhibit 5.8. Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Biological Mother/ Caregiver Race/Ethnicity

Outcome Measure (Number of Outcomes Examined)	Impact in Subgroup A (Head Start – Control) Non-Hispanic White/Other Children	Impact in Subgroup B (Head Start –Control) Non-Hispanic Black Children	Impact in Subgroup C (Head Start –Control) Hispanic Children	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE				
Direct Child Assessment (5)				
School Performance (5)				
SOCIAL-EMOTIONAL				
Parent Report (5)				
Aggressive Behavior	0.09	**-0.76*** **	-0.16	A-B** B-C**
Hyperactive Behavior		-0.47**	0.14	B-C**
Total Child Behavior Problems	-0.13	**-1.59*** **	-0.11	A-B* B-C**
Teacher Report (10)				
Closeness with Teacher	**-1.60*** **		0.18	A-C**
Child Report (4)				
Internalizing	0.04		-0.26**	A-C**
Peer Relations	**-0.32*** **		-0.02	A-C**
School	-0.08	0.31*		A-B**
HEALTH (5)				
PARENTING				
Parent Report (10)				
Supportive School Environment		-0.20	0.20**	B-C**
Parent Spanked Child in Last Week	0.04	-0.22**	0.06	A-B** B-C**
Teacher Report (2)				
Parent Participation	-0.08*		0.04	A-C*

Key:
*** $p \leq 0.01$
** $p \leq 0.05$
* $p \leq 0.10$

Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.

Exhibit 5.9. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Biological Mother/ Caregiver Race/Ethnicity

Outcome Measure (Number of Outcomes Examined)	Impact in Subgroup A (Head Start – Control) Non-Hispanic White/Other Children	Impact in Subgroup B (Head Start – Control) Non-Hispanic Black Children	Impact in Subgroup C (Head Start – Control) Hispanic Children	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE				
Direct Child Assessment (5)				
School Performance (5)				
Promotion	-0.04**	0.02		A-B**
SOCIAL-EMOTIONAL				
Parent Report (5)				
Aggressive Behavior	0.31*		-0.24	A-C*
Social Skills and Positive Approaches to Learning	0.60**		-0.12	A-C**
Teacher Report (10)				
Closeness	0.64	-1.22*		A-B*
Child Report (4)				
School		-0.17	0.22*	B-C**
HEALTH				
PARENTING				
Parent Report (10)				
Spank	0.09*		-0.10	A-C**
Teacher Report (2)				
Parent Participation	0.06**	-0.07	-0.07**	A-B** **A-C*** **

Key:
*** $p \leq 0.01$
** $p \leq 0.05$
* $p \leq 0.10$

Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.

In contrast, there was only a single within-group unfavorable impact for Black children. Impacts on Hispanic children were mixed. For Black children, teachers reported less closeness to children in the Head Start group than to children in the control group. Hispanic children showed a favorable impact on their perception of how they do in school and their enjoyment of school and there was an unfavorable impact on parents' participation in school activities as reported by teachers.

Parental Report of Depressive Symptoms

The discussion for subgroups defined by parents' reported depressive symptoms is complicated. For each age group and each variable, six difference of differences tests were conducted (each of the four levels of depressive symptoms – no, mild, moderate, and severe – compared against each of the others.) As a result, there were many significance tests conducted for this subgroup, and one would expect some of these to be significant just by chance. A further complication in presenting and discussing these findings is the fact that differential impacts can be found in any one of these six tests, and it is rare that those differential impacts are found only in tests of one group against one other. Thus, in any given subgroup, there might be several impacts all going in one direction (e.g., all favorable or all unfavorable), but the differences with other groups vary. Thus, presenting findings based on the differential impacts *across* groups rather than those *within* groups makes it rather difficult to understand the overall picture of impacts. In our presentation below, we concentrate on the direction of impacts *within* any given group, describing where the bulk of the differential impacts *across* groups were found. We note, however, that we only discuss those *within* group impacts that were first found to be statistically significant from an impact in another group. Readers should examine Exhibits 5.10 and 5.11 to understand which of the differential impacts was significant in each of these cases.

For the 4-year-old cohort, the most frequent differences in impacts occurred between children of parents reporting severe depressive symptoms and each of the other three groups, but most especially those with no or moderate symptoms. For children whose parents had severe depressive symptoms, there were several statistically significant favorable impacts on parents' reports of children's behavior, as well as a favorable impact on parents' perceptions of the effect parenting has had on their lives. Among parents reporting severe depressive symptoms, parents

in the Head Start group reported that their children were less aggressive, less hyperactive, and had fewer behavior problems than parents in the control group. Parents in the Head Start group also reported that parenting had a more positive effect on their lives than parents in the control group. In these domains, differential impacts tended to be found between this group of families with severe parental depressive symptoms and those with no or moderate symptoms, for whom there was typically no impact on these variables.

In contrast to children whose parents had severe depressive symptoms, there were several unfavorable impacts for children whose parents had mild depressive symptoms, particularly within the social emotional domain. Within this group of children whose parents had mild depressive symptoms, teachers reported both poorer relationships with and poorer behavior among children in the Head Start group than children in the control group. Teachers described Head Start children as being less likely to fall within the normal range of pro-social behavior or total behavioral difficulties and also have less closeness, more conflict, and less positive relationships with their teachers. These impacts were most commonly found to be significantly different from the impacts for children whose parents had severe depressive symptoms, for whom there were no significant impacts on teachers' reports of social emotional skills. However, the picture was not totally unfavorable for children whose parents had mild depressive symptoms. There was also a favorable impact on the reading scores of these children, as assessed by a direct child assessment, and this impact was significantly different from the lack of impact found for children whose parents reported no depressive symptoms.

The findings for children whose parents had no depressive symptoms or those with moderate depressive symptoms were more mixed in direction, with both groups showing both favorable and unfavorable impacts that were significantly different from other subgroups. For those with no depressive symptoms, there were unfavorable impacts on children's social emotional reports, with teachers reporting that Head Start children in this group showed less social competency than their control group peers, and the Head Start children themselves reported more internalizing problems than their control group counterparts. There was a mix of favorable and unfavorable impacts on parents' reports about parenting. Within the no depressive symptoms group, Head Start parents reported that parenting had a more positive effect on their lives, but also reported a more negative perception of school services than their counterparts in the control group. In most cases, these impacts for those with no depressive symptoms were

significantly different from the lack of impacts on these variables found for the severe depressive symptoms group.

For children in the moderate symptoms group, there was also a mixed picture. Head Start children in this group were more likely to be promoted than their counterparts in the control group. Yet, teachers reported that fewer of these children fell into the normal category in their behavior problems, and parents reported decreased neglectful parenting but increases in permissiveness and poorer perceptions of school services. In most cases, these impacts were significantly different from those in the no symptoms group, and some were different from the severe symptoms group as well.

For the 3-year-old cohort, the most frequent differences in impacts occurred between children of parents with moderate depressive symptoms and each of the other three groups (Exhibit 5.11). These differences spanned the social-emotional, health, and parenting domains. Among children of parents with moderate depressive symptoms, there were favorable impacts on children's reports of their relations with peers, and several favorable impacts on parenting, including parent reports of parenting styles and the impact of parenting on their lives. Within this group, Head Start participation significantly decreased the percentage of children who required care for an injury.[77]

Compared to the subgroup with moderate symptoms of depression, impacts for the other groups were less frequent. These impacts were mixed in direction. For children whose parents reported no depressive symptoms, there were favorable impacts in three direct child assessments, the PPVT, the ECLS-K Reading Assessment, and WJ III Letter Word Identification. These impacts were statistically different from the impacts for the mild and moderate symptoms groups. However, Head Start parents in the no depressive symptoms group also showed a decrease in the time they spent with their child.

[77] See Footnote 64.

Exhibit 5.10. Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Parent/Caregiver Reported Depressive Symptoms

Outcome Measure (Number of Outcomes Examined)	Impact in Subgroup A (Head Start – Control) No Symptoms	Impact in Subgroup B (Head Start – Control) Mild Symptoms	Impact in Subgroup C (Head Start – Control) Moderate Symptoms	Impact in Subgroup D (Head Start – Control) Severe Symptoms	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE					
Direct Child Assessment (5)					
ECLS-K Reading	-0.11	5.24**			A-B*
School Performance (5)					
Promotion	-0.02		0.11*		A-C*
SOCIAL-EMOTIONAL					
Parent Report (5)					
Aggressive Behavior	-0.08		0.20	-1.20**	A-D** C-D**
Hyperactive Behavior	0.09		0.26	-0.69*	A-D* C-D*
Total Behavior Problems	-0.25	-0.50	0.52	-2.21**	A-D* B-D* C-D*
Teacher Report (10)					
Pro-social Behavior+	-0.04	**-0.18*** **	0.07	0.08	A-B* B-C* B-D*
Closeness with Teacher		-1.39*		1.43	B-D**
Conflict with Teacher		2.13**		-3.11	B-D**
Positive Relationship with Teacher+		-3.40**		5.41	B-D**
Total Difficulties+		-0.15*	-0.20*	0.20	B-D** C-D*
Social Competency	-0.15*			0.38	A-D*
Child Report (5)					
Internalizing	-0.20*			0.35	A-D***

Exhibit 5.10. Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Parent/Caregiver Reported Depressive Symptoms (continued)

Outcome Measure (Number of Outcomes Examined)	Impact in Subgroup A (Head Start – Control) No Symptoms	Impact in Subgroup B (Head Start – Control) Mild Symptoms	Impact in Subgroup C (Head Start – Control) Moderate Symptoms	Impact in Subgroup D (Head Start – Control) Severe Symptoms	Statistically Significant Differences in Impacts Between Subgroups
HEALTH (5)					
PARENTING					
Parent Report (10)					
Parenting Style: Neglectful	0.03	0.05	**-0.24*** **		**A-C*** ** **B-C*** **
Parenting Style: Permissive	-0.05		0.15*		A-C**
Effect of Parenting on Parent's Life	0.14*		-0.19	0.47*	A-C* C-D**
Parent Perception of School Services	-0.21*		-0.28*	0.17	A-D* C-D**
Teacher Report (2)					

Key:
*** $p \leq 0.01$
** $p \leq 0.05$
* $p \leq 0.10$

Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.

NOTE: + indicates scores from the Strengths and Difficulties Questionnaire that are reported as the percentage difference between Head Start and control group children in the normal category. The teacher- reported scores include Conduct Problems, Emotional Symptoms, Hyperactivity, Peer Problems, Pro-social Behavior, and Total Difficulties.

Exhibit 5.11. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Parent/Caregiver Reported Depressive Symptoms

Outcome Measure (Number of Outcomes Examined)	Impact in Subgroup A (Head Start – Control) No Symptoms	Impact in Subgroup B (Head Start – Control) Mild Symptoms	Impact in Subgroup C (Head Start – Control) Moderate Symptoms	Impact in Subgroup D (Head Start – Control) Severe Symptoms	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE					
Direct Child Assessment (5)					
PPVT	4.48**		-4.05		A-C*
ECLS-K Reading	3.13**	-2.59			A-B*
WJIII Letter-Word Identification	5.79*	-3.89	-7.48		A-B** A-C**
School Performance (5)					
SOCIAL-EMOTIONAL					
Parent Report (5)					
Teacher Report (10)					
Emotional Symptoms +	0.03	0.07	-0.01	-0.15**	A-D*** B-D*** C-D**
Pro-social+		**-0.18*** **	0.08		B-C*
Positive Relationships with Teacher	-0.32	**-3.20*** **			A-B*
Child Report (4)					
Peer Relations	0.05	-0.07	0.58***	0.01	A-C** B-C** C-D**
HEALTH (5)					
Child Had Care for Injury in Last Month	-0.02	0.00	-0.15**		A-C* B-C*

Exhibit 5.11. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Parent/Caregiver Reported Depressive Symptoms (continued)

Outcome Measure (Number of Outcomes Examined)	Impact in Subgroup A (Head Start – Control) No Symptoms	Impact in Subgroup B (Head Start – Control) Mild Symptoms	Impact in Subgroup C (Head Start – Control) Moderate Symptoms	Impact in Subgroup D (Head Start – Control) Severe Symptoms	Statistically Significant Differences in Impacts Between Subgroups
PARENTING					
Parent Report (10)					
Supportive School Environment	-0.01	-0.03	-0.08	0.35**	A-D** B-D* C-D**
Parenting Style: Authoritarian	-0.01	0.01	**-0.19*** **	0.04	A-C** B-C** C-D**
Parenting Style: Authoritative			**0.27*** **	-0.01	C-D*
Effect of Parenting on Parent's Life	-0.13		0.25*		A-C*
Time Spent With Child	-0.22***		0.17	0.30	A-C* A-D**
Teacher Report (2)					

Key:
*** $p \leq 0.01$
** $p \leq 0.05$
* $p \leq 0.10$

Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.

NOTE: + indicates scores from the Strengths and Difficulties Questionnaire that are reported as the percentage difference between Head Start and control group children in the normal category. The teacher-reported scores include Conduct Problems, Emotional Symptoms, Hyperactivity, Peer Problems, Pro-social Behavior, and Total Difficulties.

For both the mild and the severe symptoms groups, there were only two statistically significant impacts within these groups that were also significantly different from impacts in other groups. In the mild group, these were both unfavorable, with teachers reporting that Head Start children in this group were less likely to show normal pro-social behavior and had poorer relationships with their teachers than their counterparts in the control group. In the severe symptoms group, these impacts were mixed, with an unfavorable impact on emotional symptoms and a favorable impact on parents' perceptions of the supportiveness of the school environment.

Household Risk

For the 4-year-old cohort, subgroups defined by level of household risk showed few findings that were mixed in direction (see Exhibit 5.12.) There were favorable impacts on two teacher-reported measures of reading and language arts ability and skills for children in high risk households, but unfavorable impacts on parents' reports of doing activities with their children. For children in moderate risk households, there were two unfavorable impacts: one on children's reports of their relationships with their peers and one on whether the child has health insurance according to his or her parents. In most cases these impacts were significantly different from those for the low/no household risk group, for whom none of these impacts was significant.

Within the 3-year-old cohort, differences were typically found between children in high risk households and those from the other two groups, with several favorable impacts for children in high risk households (Exhibit 5.13.) For children in high risk households, there were favorable impacts on the ECLS-K Reading and WJIII Letter-Word Identification assessments. There were also favorable impacts on teachers' reports of the reading and language arts skills of children in high risk households and on parents' report of their parenting styles and activities with their children.

In contrast to impacts for high risk households, impacts on moderate risk households were unfavorable, and impacts on those with low/no household risk were more mixed in direction. For moderate risk households, unfavorable impacts were found in the cognitive and parenting domains. Teachers' report of math skills showed an unfavorable impact of Head Start for children from moderate risk households. While moderate risk household Head Start parents reported less authoritarian parenting styles than their control group peers, they also reported

Exhibit 5.12. Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Household Risk Index

Outcome Measure (Number of Outcomes Examined)	Impact in Subgroup A (Head Start – Control) Low/No Household Risk	Impact in Subgroup B (Head Start –Control) Moderate Household Risk	Impact in Subgroup C (Head Start –Control) High Household Risk	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE				
Direct Child Assessment (5)				
School Performance (5)				
Reading/Language Arts Ability	0.01		0.30*	A-C*
Reading/Language Arts Skills	-0.09		0.69*	A-C**
SOCIAL-EMOTIONAL				
Parent Report (5)				
Teacher Report (10)				
Child Report (4)				
Peer Relations	-0.09	-0.49**	0.35	A-B* B-C**
HEALTH (5)				
Child Has Health Insurance Coverage	0.01	**-0.08****		A-B*
PARENTING				
Parent Report (10)				
Doing Things Together	0.03	-0.01	-0.62**	A-C** B-C**
Teacher Report (2)				

Key:
*** $p \leq 0.01$
** $p \leq 0.05$
* $p \leq 0.10$

Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.

Exhibit 5.13. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Household Risk Index

Outcome Measure (Number of Outcomes Examined)	Impact in Subgroup A (Head Start – Control) Low/No Household Risk	Impact in Subgroup B (Head Start –Control) Moderate Household Risk	Impact in Subgroup C (Head Start –Control) High Household Risk	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE				
Direct Child Assessment (5)				
ECLS-K Reading	-0.16	-3.75	**8.85*** **	**A-C*** ** **B-C*** **
WJIII Letter-Word Identification	0.71	-6.55	**15.82*** **	**A-C*** ** **B-C*** **
School Performance (5)				
Math Skills	0.05	-0.31**	0.26	A-B** B-C*
Reading/Language Arts Skills	-0.05	-0.26	0.40*	A-C* **B-C*** **
SOCIAL-EMOTIONAL				
Parent Report (5)				
Teacher Report (10)				
Pro-social Behavior+	-0.08**	0.06		A-B**
Child Report (4)				
HEALTH (5)				
Child Needs Ongoing Care	-0.01	**0.18*** **	-0.06	**A-B*** ** B-C**

Exhibit 5.13. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Household Risk Index (continued)

Outcome Measure (Number of Outcomes Examined)	Impact in Subgroup A (Head Start – Control) Low/No Household Risk	Impact in Subgroup B (Head Start –Control) Moderate Household Risk	Impact in Subgroup C (Head Start –Control) High Household Risk	Statistically Significant Differences in Impacts Between Subgroups
PARENTING				
Parent Report (10)				
Doing Things Together		-0.11	0.36**	B-C*
Parenting Style: Authoritarian	0.01	-0.14**		**A-B****
Parenting Style: Authoritative	0.07*		0.26**	A-C*
Parenting Style: Permissive	-0.06*	0.11*	-0.19*	**A-B**** B-C**
Effect of Parenting on Parent's Life	-0.11*	0.21		A-B**
Teacher Report (2)				

Key:

***p ≤ 0.01

**p ≤ 0.05

*p ≤ 0.10

Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.

NOTE: + indicates scores from the Strengths and Difficulties Questionnaire that are reported as the percentage difference between Head Start and control group children in the normal category. The teacher-reported scores include Conduct Problems, Emotional Symptoms, Hyperactivity, Peer Problems, Pro-social Behavior, and Total Difficulties.

more permissive parenting styles. Within moderate risk households, there was also an impact on the child's needing ongoing care. For those with no or low household risk, there were mixed impacts, with a reduction in the percentage of children whose teachers reported they had normal pro-social behavior, but a decrease in permissive and an increase in authoritative parenting styles.

Urbanicity

Among subgroups defined by urbanicity, the majority of the impacts for the 4-year-old cohort were found for children from the non-urban areas, and the impacts were mixed (Exhibit 5.14). For children in non-urban areas, teachers' reports of their closeness and relationships with children showed unfavorable impacts of Head Start. Non-urban Head Start children's reports of peer relations and externalizing (anger and distractibility) showed unfavorable impacts of Head Start. Finally, non-urban Head Start parents' reports of the effects of parenting on their lives showed favorable impacts of Head Start.

Teacher reports showed one unfavorable impact on Head Start children from urban areas. They reported that Head Start children were more likely to exhibit emotional symptoms than control group children.

There were also mixed effects for the 3-year-old cohort (Exhibit 5.15). By the end of 3rd grade, teacher reports showed an unfavorable impact for non-urban children in their pro-social behaviors, but a favorable impact on total difficulties. Additionally, non-urban children showed favorable impacts on their reports of school experiences. Contrary to the findings for the 4-year-old cohort, non-urban Head Start parents' reports showed an unfavorable impact on the effect of parenting on their lives.

Exhibit 5.14. Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Urbanicity

Outcome Measure (Number of Outcomes Examined)	Impact in Subgroup A (Head Start – Control) Not Urban	Impact in Subgroup B (Head Start – Control) Urban	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE			
Direct Child Assessment (5)			
School Performance (5)			
SOCIAL-EMOTIONAL			
Parent Report (5)			
Teacher Report (10)			
Emotional Symptoms+	0.01	**-0.07*** **	A-B**
Closeness with Teacher	**-2.42*** **	-0.29	A-B**
Positive Relationship with Teacher	**-4.28*** **	-0.73	A-B*
Child Report (4)			
Externalizing	**-0.45*** **	-0.03	A-B**
Peer Relations	**-0.40** **	-0.09	A-B*
HEALTH (5)			
Dental Care	0.09**	0.00	A-B*
PARENTING			
Parent Report (10)			
Effect of Parenting on Parent's Life	**0.38*** **	-0.01	A-B**
Teacher Report (2)			

Key:
*******$p \leq 0.01$
******$p \leq 0.05$
*****$p \leq 0.10$

Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.

NOTE: + indicates scores from the Strengths and Difficulties Questionnaire that are reported as the percentage difference between Head Start and control group children in the normal category. The teacher-reported scores include Conduct Problems, Emotional Symptoms, Hyperactivity, Peer Problems, Pro-social Behavior, and Total Difficulties.

Exhibit 5.15. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Urbanicity

Outcome Measure (Number of Outcomes Examined)	Impact in Subgroup A (Head Start – Control) Not Urban	Impact in Subgroup B (Head Start – Control) Urban	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE			
Direct Child Assessment (5)			
School Performance (5)			
SOCIAL-EMOTIONAL			
Parent Report (5)			
Teacher Report (10)			
Pro-social Behavior+	-0.15**	-0.02	A-B*
Total Difficulties+	0.11*	-0.02	A-B*
Child Report (4)			
School	0.23**	0.00	A-B*
HEALTH (5)			
PARENTING			
Parent Report (10)			
Effect of Parenting on Parent's Life	-0.31**	0.03	A-B**
Teacher Report (2)			

Key:
***$p \leq 0.01$
**$p \leq 0.05$
*$p \leq 0.10$

Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.

NOTE: + indicates scores from the Strengths and Difficulties Questionnaire that are reported as the percentage difference between Head Start and control group children in the normal category. The teacher-reported scores include Conduct Problems, Emotional Symptoms, Hyperactivity, Peer Problems, Pro-social Behavior, and Total Difficulties.

Impacts That Are Sustained Through 3rd Grade

Below we describe subgroup patterns over time, addressing three different patterns of interest: (1) 3rd grade impacts that continued patterns that occurred through the end of 1st grade; (2) new patterns of impacts that were found in 3rd grade; and (3) patterns that were demonstrated through the end of 1st grade but not sustained through the end of 3rd grade. Exhibit 5.16 (4-year-old cohort) and Exhibit 5.17 (3-year-old cohort) summarize the number of impacts in each domain from the Head Start years through 3rd grade. In addition, all the subgroup findings from

Exhibit 5.16. Summary of the Subgroup Findings for the 4-Year-Old Cohort

Dimensions	Subgroup	Year	Cognitive			Social-Emotional			Health	Parenting	
			Language and Literacy	Math Skills	School Performance	Parent Report	Teacher Report	Child Report	Parent Report	Parent Report	Teacher Report
Child's Pre-Academic Skills	In Lowest Quartile	HS			NA	2/9	NA	NA			NA
		K	1/9	1/3			4/11	NA		1/9 1/9*	
		1			1/5		2/11	NA		1/8	
		3								1/10 2/10	
	Not in Lowest Quartile	HS			NA		NA	NA			NA
		K						NA			
		1		1/4				NA			
		3						1/4		1/10	
Child's Home Language	English Speaking	HS			NA		NA	NA			NA
		K						NA			
		1						NA			
		3									
	Dual Language Learners	HS	1/9		NA		NA	NA	1/5 1/5	1/5	NA
		K			1/5			NA	1/5		
		1						NA	1/5	1/8	
		3								1/10	
Special Needs	Special Needs	HS			NA		NA	NA			NA
		K				1/9	1/11	NA	1/5		
		1						NA	1/5*		
		3									
	Not Special Needs	HS			NA		NA	NA			NA
		K					1/11	NA			
		1						NA			
		3	1/3								

Exhibit 5.16. Summary of the Subgroup Findings for the 4-Year-Old Cohort (continued)

Dimensions	Subgroup	Year	Cognitive			Social-Emotional			Health	Parenting	
			Language and Literacy	Math Skills	School Performance	Parent Report	Teacher Report	Child Report	Parent Report	Parent Report	Teacher Report
Biological Mother Race/ Ethnicity	Non-Hispanic White/ Other	HS			NA		NA	NA	1/5		NA
		K	1/9				1/11	NA		1/9 1/9	
		1					3/11	NA			
		3					1/10	1/4			1/2
	Non-Hispanic Black	HS	1/9		NA		NA	NA			NA
		K	2/9				5/11	NA			
		1						NA	1/5	2/8	
		3				3/5		1/4		1/10	
	Hispanic	HS	1/9	1/2	NA		NA	NA			NA
		K						NA		1/9	
		1				1/9		NA		1/8	
		3						1/4		1/10	
Parent Depressive Symptoms	No Symptoms	HS	1/9		NA		NA	NA			NA
		K				1/9		NA		1/9	
		1						NA			
		3					1/10	1/4		1/10 1/10	
	Mild Symptoms	HS	3/9	1/2	NA	2/9	NA	NA			NA
		K				1/9	1/11	NA			
		1	3/11	1/4		1/9		NA			
		3	1/3				5/10				
	Moderate Symptoms	HS	1/9		NA		NA	NA			NA
		K						NA	1/5*		1/2
		1					1/11	NA			1/2
		3			1/5		1/10			1/10* 2/10	
	Severe Symptoms	HS	3/9		NA		NA	NA	1/5*		NA
		K				1/9		NA	1/5	1/9	
		1		1/4		1/9	1/11	NA	2/5*	1/8	
		3				3/5				1/10	

Exhibit 5.16. Summary of the Subgroup Findings for the 4-Year-Old Cohort (continued)

Dimensions	Subgroup	Year	Cognitive			Social-Emotional			Health	Parenting	
			Language and Literacy	Math Skills	School Performance	Parent Report	Teacher Report	Child Report	Parent Report	Parent Report	Teacher Report
Household Risk	Low/No Risk	HS	2/9		NA	1/9	NA	NA	1/5 1/5		N/A
		K						NA		1/9	
		1					1/11	NA		1/8	
		3									
	Moderate Risk	HS	1/9	1/2	NA		NA	NA			NA
		K					1/11 1/11	NA			
		1	1/11					NA	1/5	1/8	
		3						1/4	1/5		
	High Risk	HS			NA	2/9	NA	NA	1/5		NA
		K				1/9	2/11	NA			
		1					1/11	NA	1/5*		
		3			2/5					1/10	
Urbanicity	Urban	HS	1/9	1/2	NA		NA	NA	1/5		NA
		K						NA			
		1						NA			
		3					1/10				
	Not Urban	HS	1/9		NA		NA	NA	1/5	1/5	NA
		K	1/9			1/9		NA			1/2
		1			1/5			NA			1/2
		3					2/10	2/4	1/5	1/10	

KEY:

Numeral indicates the number of significant outcomes out of the total number of outcomes for that cell.

Dark gray cell indicates a significant favorable impact ($p \leq 0.10$).

Diagonal mark cell indicates a significant unfavorable impact ($p \leq 0.10$).

Blank cell indicates no significant impact.

An * indicates that the interpretation of the outcome is unclear.

NA indicates that no data were collected for this outcome at this data collection point.

Exhibit 5.17. Summary of the Subgroup Findings for the 3-Year-Old Cohort

Dimensions	Subgroup	Year	Cognitive			Social-Emotional			Health	Parenting	
			Language and Literacy	Math Skills	School Performance	Parent Report	Teacher Report	Child Report	Parent Report	Parent Report	Teacher Report
Child's Pre-Academic Skills	In Lowest Quartile	HS			NA	1/9	NA	NA			NA
		Age 4	4/9	1/2	NA	1/9	NA	NA		2/9	NA
		K						NA			
		1						NA			
		3									
	Not in Lowest Quartile	HS			NA		NA	NA		1/5*	NA
		Age 4			NA		NA	NA			NA
		K					2/11	NA			
		1						NA			
		3				1/5					
Child's Home Language	English Speaking	HS	1/9		NA	1/9	NA	NA	1/5	1/5	NA
		Age 4			NA	1/9	NA	NA			NA
		K		2/3				NA			
		1						NA		1/8	
		3				1/5			1/5		
	Dual Language Learners	HS	2/9		NA		NA	NA	2/5		NA
		Age 4	5/9	1/2	NA	1/9	NA	NA			NA
		K	2/9					NA	1/5*		
		1						NA			1/2
		3							1/5		
Special Needs	Special Needs	HS			NA		NA	NA		1/5*	NA
		Age 4			NA		NA	NA	1/5*	1/9	NA
		K						NA		1/9 2/9	
		1		4/4			4/11	NA		1/8*	
		3									
	Not Special Needs	HS			NA		NA	NA			NA
		Age 4			NA		NA	NA			NA
		K						NA			
		1						NA			
		3									

Exhibit 5.17. Summary of the Subgroup Findings for the 3-Year-Old Cohort (continued)

Dimensions	Subgroup	Year	Cognitive			Social-Emotional			Health	Parenting	
			Language and Literacy	Math Skills	School Performance	Parent Report	Teacher Report	Child Report	Parent Report	Parent Report	Teacher Report
Biological Mother Race Ethnicity	Non-Hispanic White/ Other	HS	1/9		NA	3/9	NA	NA	1/5 1/5*		NA
		Age 4	1/9		NA	1/9	NA	NA			NA
		K	1/9			1/9		NA	1/5*	1/9	
		1	1/11			1/9 2/9		NA	1/5*		
		3			1/5	1/5 1/5				1/10	1/2
	Non-Hispanic Black	HS	1/9		NA	1/9	NA	NA	1/5		NA
		Age 4			NA		NA	NA			NA
		K	1/9	1/3				NA	1/5	2/9 1/9*	
		1				4/9		NA		1/8 1/8	
		3					1/10				
	Hispanic	HS	1/9		NA		NA	NA	1/5 1/5*		NA
		Age 4	2/9	1/2	NA		NA	NA			NA
		K	1/9				1/11	NA		3/9	
		1		1/4		1/9 1/9	1/11	NA			1/2
		3						1/4			1/2
Parent Depressive Symptoms	No Symptoms	HS	1/9		NA	1/9	NA	NA		1/5	NA
		Age 4	5/9	1/2	NA	2/9	NA	NA		1/9	NA
		K	1/9			5/9		NA		1/9	
		1	7/11	2/4			1/11	NA		1/8	
		3	3/3							1/10	
	Mild Symptoms	HS			NA	1/9	NA	NA			NA
		Age 4	1/9		NA		NA	NA			NA
		K				1/9	3/11	NA		1/9	
		1			3/5		3/11	NA			
		3					2/10				
	Moderate Symptoms	HS			NA		NA	NA		1/5	NA
		Age 4	4/9		NA	1/9	NA	NA	2/5		NA
		K			2/5	1/9	5/11	NA		1/9 1/9 1/9*	1/2
		1	9/11		1/5		2/11	NA	1/5*		
		3						1/4	1/5*	3/10	
	Severe Symptoms	HS	1/9		NA	1/9	NA	NA	2/5		NA
		Age 4	1/9		NA	1/9	NA	NA			NA
		K			1/5			NA			
		1					2/11	NA	1/5*		
		3					1/10			1/10	

Exhibit 5.17. Summary of the Subgroup Findings for the 3-Year-Old Cohort (continued)

Dimensions	Subgroup	Year	Cognitive			Social-Emotional			Health	Parenting	
			Language and Literacy	Math Skills	School Performance	Parent Report	Teacher Report	Child Report	Parent Report	Parent Report	Teacher Report
Household Risk	Low/No Risk	HS	2/9	1/2	NA		NA	NA	1/5		NA
		Age 4	3/9		NA		NA	NA			NA
		K						NA			
		1						NA			
		3					1/10			2/10	1/10
	Moderate Risk	HS			NA	1/9	NA	NA	1/5*		NA
		Age 4			NA	1/9	NA	NA	1/5*		NA
		K	1/9	2/3				NA			
		1	1/11	1/4		3/9	1/11	NA		1/8	
		3			1/5				1/5*	1/10	1/10
	High Risk	HS	2/9		NA		NA	NA	1/5		NA
		Age 4	1/9		NA		NA	NA	1/5*		NA
		K	1/9				2/11	NA			
		1	5/11					NA			
		3	2/3		1/5					3/10	
Urbanicity	Urban	HS	3/9		NA	1/9	NA	NA	1/5*		NA
		Age 4			NA		NA	NA			NA
		K						NA			
		1						NA			
		3									
	Not Urban	HS	4/9		NA	3/9	NA	NA	1/5*	1/5	NA
		Age 4		1/2	NA	1/9	NA	NA	1/5		NA
		K	1/9					NA	1/5*		
		1	6/11	1/4				NA			
		3					2/10	1/4		1/10	

KEY:

Numeral indicates the number of significant outcomes out of the total number of outcomes for that cell.

Dark gray cell indicates a significant favorable impact ($p \leq 0.10$).

Diagonal mark cell indicates a significant unfavorable impact ($p \leq 0.10$).

Blank cell indicates no significant impact.

An * indicates that the interpretation of the outcome is unclear.

NA indicates that no data were collected for this outcome at this data collection point.

the Head Start years through 3rd grade that met the criteria described on page 1 in this chapter, are presented in Appendix E.

Child's Pre-Academic Skills. For the 4-year-old cohort, children in the lowest academic quartile at baseline showed benefits of Head Start in the social-emotional domain through 1st grade, but this pattern was not sustained at the end of 3rd grade. At the end of the Head Start year, there were favorable impacts for the lowest quartile children compared to their non-lowest quartile counterparts on parents' report of their relationship with the child. In the early school years (K and 1st grade), teacher reports showed favorable impacts for lowest quartile children on oppositional behavior, problems with peer interaction, closeness, conflict, and positive relationships with the teacher than for non-lowest quartile children. By the end of 3rd grade, there were no impacts reported in the social-emotional domain for children in either subgroup. The impacts for parenting practices were mixed for children in the lowest quartile subgroup and this continued through 3rd grade.

Over the years of the study, there were negative social-emotional impacts for the 3-year-old cohort children not in the lowest quartile, although for differing outcomes and from different reporters. In kindergarten, teachers reported more aggressive behavior and peer problems for Head Start children, and at the end of 3rd grade, parents reported that Head Start children not in the lowest quartile were more likely to be withdrawn than their counterparts.

Children with Special Needs. For the 4-year-old cohort, there is no clear pattern of findings for the special needs subgroups. For the 3-year-old cohort, children with special needs benefitted from Head Start in the cognitive and social-emotional domains at the end of 1st grade, but the impacts were not sustained through the end of 3rd grade.

Dual Language Learners. Prior to 3rd grade, there is some evidence that Head Start Dual Language Learners in the 4-year-old cohort benefited more in the health domain than their counterparts, but this pattern was not sustained in 3rd grade. However, in the parenting practices domain, favorable impacts were sustained through the end of 3rd grade, with Head Start group parents reporting a higher use of safety practices at the end of the Head Start year, a reduction in neglectful parenting style at the end of 1st grade, and a more supportive school environment at the end of 3rd grade.

For the 3-year-old cohort, the favorable cognitive impacts found earlier for dual language learners were not found at the end of 3rd grade. However, there was on an ongoing favorable impact in the health domain on child's health status at the end of the Head Start year and the end of 3rd grade. Also, the earlier parent-reported favorable social-emotional impacts for English-speaking children in the Head Start group were found at the end of 3rd grade and at the end of the Head Start and age 4 years.

Race/Ethnicity. Head Start Black children in the 4-year-old cohort continued to show favorable impacts over their counterparts. For children in the 4-year old cohort, Black children experienced favorable impacts in the social-emotional domain at the end of kindergarten as reported by teachers, including reduced inattentiveness; fewer problems with structured learning, peer interactions, or teacher interactions; and better relationships with teachers. This pattern continued at the end of 3rd grade with parents of Black Head Start group children reporting less aggressive, hyperactive, and total problem behaviors and children themselves reporting that they did better in school and enjoyed school more than the children in the control group. The majority of these impacts were significantly different from both White and Hispanic children. There was also evidence of sustained favorable impacts in the parenting practices domain, with Black parents of Head Start group children reporting spanking their children less at the end of 1st and 3rd grades and a reduction in the permissive parenting style at the end of 1st grade relative to the parents of the control group children.

For Hispanic children in the 4-year-old cohort, favorable impacts were found at the end of kindergarten, 1st and 3rd grades in the parenting domain, with parents of Hispanic Head Start group children reporting more cultural enrichment, less neglectful parenting style, and a more supportive school environment than Hispanic parents of control group children.

In contrast to a more positive picture for the Head Start 4-year-old cohort of Black children, White children were more likely to experience unfavorable impacts reported by teachers at the end of 1st grade, particularly in the social emotional domain, and this pattern continued through 3rd grade.

In contrast to the 4-year-old cohort, there is no clear pattern of benefit for one racial or ethnic subgroup in the 3-year-old cohort. For Black children in the 3-year-old cohort, favorable impacts were found at the end of 1st grade in the social-emotional domain, but the

pattern was not maintained at the end of 3rd grade. At the end of 3rd grade, there was one unfavorable impact with teachers reporting less closeness with Black children in the Head Start group than with Black children in the control group.

For the 3-year-old Hispanic cohort, earlier mixed impacts were found for Head Start children in the social-emotional domain as reported by both parents and teachers. By the end of 3rd grade, a favorable impact in this domain was reported by children. Hispanic Head Start group children reported that they liked school and enjoyed school more than children in the control group.

Mixed impacts were found through the end of 1st grade and continued through the end of 3rd grade for White children in the 3-year-old cohort in the social-emotional and parenting domains.

Parental Report of Depressive Symptoms. For the 4-year-old cohort, children of parents with mild depressive symptoms experienced favorable cognitive impacts through the end of 1st grade, and this pattern was repeated in 3rd grade. At the end of the Head Start year, this subgroup experienced several benefits of Head Start in language and literacy compared to children of parents with other levels of reported depressive symptoms. No cognitive impacts were found in kindergarten, but the language and literacy impacts re-appeared at the end of 1st grade and were sustained through the end of 3rd grade with favorable impacts on the ECLS-K Reading Assessment.

As reported earlier, for the 4-year-old cohort, the most frequent differences in impacts in 3rd grade occurred between children of parents reporting severe depressive symptoms and each of the other three groups. Specifically, for children of parents with reported severe depressive symptoms predominately favorable impacts were sustained in the parenting and social-emotional domains.

For the 3-year-old cohort in previous years, there was a pattern that Head Start had favorable impacts through 1st grade in the cognitive, social-emotional, and parenting domains for children of parents with no reported depressive symptoms compared to other children. This pattern was sustained only in the cognitive domain at the end of 3rd grade. At the end of 3rd grade, children of parents with no reported depressive symptoms benefited from Head Start on all three assessments of language and literacy.

For children of parents with mild depressive symptoms, the pattern of mostly unfavorable impacts continued through the end of 3rd grade in the social-emotional domain, including teacher report of poorer relationships with children and poorer child behavior. The previously teacher-reported unfavorable social-emotional impacts for children of parents with moderate symptoms were not seen at the end of 3rd grade. Instead, Head Start children of parents with moderate depressive symptoms reported more favorable peer relations than any of the other depressive symptoms subgroups.

Household Risk. No pattern of sustained impacts was found for the 4-year-old cohort of children from no risk or high risk households. Sustained unfavorable impacts in the health domain were found for children from moderate risk households. Over time, mixed impacts were found in the social-emotional domain for children from the moderate risk households.

For the 3-year-old cohort, as found in 1st grade, children from high-risk households showed sustained favorable cognitive impacts through the end of 3rd grade. Children from high-risk households experienced benefits in five direct assessments of academic skills at the end of 1st grade. At the end of 3rd grade, there were favorable impacts on two of the three language and literacy assessments and the teacher's rating of the child's reading/language arts skills. In contrast, favorable cognitive impacts for children from low/no risk households through the end of the age 4 year became neutral during the school years, whereas impacts for the moderate household risk subgroup moved from neutral in the first two years to unfavorable in kindergarten, 1st and 3rd grades. No pattern was apparent in the other domains for the 3-year-old cohort in the household risk subgroups.

Urbanicity. As with previous years, there were few impacts in any of the domains for the 4-year-old cohort from urban areas. Most of the impacts for children from non-urban areas were unfavorable in the cognitive and social-emotional domains. In kindergarten, Head Start children had unfavorable impacts in language and literacy skills and teacher report of school performance in 1st grade. Unfavorable social-emotional impacts were reported by parents in kindergarten and then by teacher and child reports in 3rd grade. However, in the health domain, there was a favorable impact on receipt of dental care for non-urban children at the end of the Head Start year and 3rd grade. Also, parents of Head Start children from non-urban areas

reported less spanking at the end of kindergarten and a favorable impact of parenting on their life at the end of 3rd grade.

Although children in the 3-year-old cohort from non-urban settings showed long-term cognitive benefits from Head Start through the end of 1st grade, these benefits were not sustained through the end of 3rd grade. For the 3-year-old cohort, there were several favorable impacts in the social emotional domain for the non-urban children through the age 4 year. By the end of 3rd grade, the impacts were mixed. At the end of 3rd grade, teachers reported more Head Start children from non-urban settings in the normal category for total behavioral difficulties, yet they reported fewer Head Start children in the normal category for pro-social behavior as compared to children in the control group. Non-urban children themselves were more likely to report that they liked and enjoyed school in the 3rd grade than their control group counterparts.

Summary of Findings

This chapter reviewed the evidence of differential impacts for key subgroups of children at the end of 3rd grade. We then looked for evidence of patterns of impacts for each subgroup from pre-K through 3rd grade to get a better sense of who benefits most from Head Start.

At the end of 3rd grade, the most striking subgroup finding was related to children from high risk households. For this subgroup, children in the 3-year old cohort maintained sustained cognitive impacts across all the years from pre-K through 3rd grade. At the end of 3rd grade, the Head Start children from high risk households showed favorable impacts as compared to children in low/no risk and moderate risk households on the ECLS-K Reading Assessment, the WJIII Letter-Word Identification, and the teacher-reported reading/language arts skills. These children who started out with more severe challenges than their peers were found to have multiple positive impacts on the direct student assessments over time. For the 4-year-old cohort, Head Start children of parents reporting mild depressive symptoms also demonstrated favorable cognitive impacts through the end of 3rd grade.

Several of the subgroups experienced sustained benefits of Head Start in one or all domains and for multiple outcome measures. These subgroups were found in both age cohorts. Among the 4-year-olds, the subgroups that stood out were children of parents with reported

severe depressive symptoms and Black children. Among the 3-year-old cohort, the most notable subgroups with positive impacts were children of parents with no reported depressive symptoms and children from high risk households. Many other subgroups for both age cohorts experienced a mixture of favorable and unfavorable impacts, particularly in the social-emotional domain.

Finally, we identified several subgroups that experienced solely–or primarily–unfavorable impacts of Head Start. These included for the 4-year-old cohort, White children, with unfavorable impacts in the social-emotional domain, and for the 3-year-old cohort, children of parents with mild depressive symptoms, with unfavorable impacts in the social-emotional domain.

Chapter 6: Conclusions

Introduction

Head Start seeks to improve educational and developmental outcomes for children from severely economically disadvantaged families. As stated in the *Head Start Act,*[78] the intent of the program is to "promote the school readiness of low-income children by enhancing their cognitive, social, and emotional development (1) in a learning environment that supports children's growth in language, literacy, mathematics, science, social and emotional functioning, creative arts, physical skills, and approaches to learning and (2) through the provision to low-income children and their families of health, educational, nutritional, social, and other services that are determined, based on family needs assessments, to be necessary." {Sec. 645. [42 U.S.C. 9840] (a) (1)}

The goals of Head Start are supported by research that documents the importance of early skill development in children. For example, Cunha *et al*. (2006) noted that early investment in a child's skill development facilitates the productivity of investments later in life. Similarly, Heckman (2008) argued that intervention early in the life of a disadvantaged child has better returns than later investment. Another study examined six longitudinal datasets and determined that early math, reading, and attention skills were strong predictors of later achievement (Duncan et al., 2007).

This chapter integrates the results discussed in this report as well as the findings through 1st grade from the *Head Start Impact Study*[79] to provide an overall cohesive discussion about what we have learned regarding the extent to which Head Start meets these stated goals. The narrative begins with the context within which impacts on child and family outcomes are expected to occur through an assessment of how the availability of Head Start shapes the preschool and early elementary school experiences of newly entering 3- and 4-year-old eligible children. We then examine the extent to which Head Start supports children's school readiness by looking at its impact on children's cognitive, social-emotional, and health outcomes and on

[78] Public Law 110-134 *Improving Head Start for School Readiness Act of 2007*, signed into law on December 12, 2007.

[79] U.S. Department of Health and Human Services, Administration for Children and Families (January, 2010). *Head Start Impact Study: Final Report*, Washington, DC. http://www.acf.hhs.gov/programs/opre/hs/impact_study/reports/impact_study/hs_impact_study_final.pdf.

the parenting practices of their primary caregivers from preschool through 3rd grade. Next, we examine how impacts varied by child and family characteristics through the end of 3rd grade. Finally, we discuss the broader themes and implications of this set of findings.

How Does Gaining Access to Head Start Affect Children's Head Start and Early Elementary School Experiences?

Providing access to Head Start had a positive impact on children's preschool experiences but little difference on children's early elementary school experience. There were statistically significant differences between the Head Start group and the control group on every measure of children's preschool experiences measured in this study. These effects were found both for the 4-year-old cohort and for the 3-year-old cohort during the year they were admitted to Head Start. The differences in magnitude were quite large, driven in part by the large proportion of children in the control group who were in parent care (i.e., nearly four out of ten children remained at home with their parents when Head Start was unavailable to them). Yet, analyses excluding those children, and thus comparing only children in the Head Start and control groups who were in non-parental care, largely showed the same pattern of differences, albeit somewhat smaller.

The preschool experiences of children in the 3-year-old cohort were very different in the second year of the study, when most were 4 years old. The majority of the children (both Head Start and control group) were in some type of center-based care by this year, and with three small exceptions, the observed treatment and control differences disappeared in the age 4 year. That is, once the control group had access to Head Start, the earlier differences on measures of their early childhood care environments all but vanished.

While on average, access to Head Start resulted in more positive experiences for children, not all children in the Head Start group had the same quality of experience. The majority (70 percent) of Head Start children in both cohorts were in centers with overall average ECERS-R scores of at least a five on a seven-point scale, indicating a good or better environment. While most children (60 percent) were also in classrooms that emphasized language and literacy and math activities; about 30 percent of the Head Start children had teachers with a BA degree; and slightly more than one-third had teachers who had received 25

hours or more of training in the last year, there were also sizable groups of children who did not experience these quality factors. Thus, the nature and quality of the experience varied—for some children it was very good, while for other children it was less so. Both the average high quality and the variation may be important in understanding impacts on child and family outcomes.

Providing access to Head Start did not have much impact on the types of schools children attended from kindergarten through 3rd grade. There were few systematic differences between the control and Head Start groups in the type or quality of schools that children attended. Where differences were found, they did not paint a clear picture of better or worse schools for the Head Start group or the control group.

The vast majority of children in both the Head Start and the control group attended public elementary schools. These schools tended to fall in the middle of their respective states on their statewide reading and math assessments, indicating that-on average-the schools attended by the study children were not among the worst or best schools in their respective states. Not surprisingly, the study children attended schools with much higher levels of poverty than schools nationwide and schools with higher proportions of minority students.

What Is the Impact of Head Start on Children and Families?

Language and Literacy Development. There is clear evidence that Head Start had an impact on children's language and literacy development while children were in Head Start. These effects, albeit modest in magnitude, were found for both age cohorts during their first year of admission to the Head Start program. However, these early effects were no longer evident in elementary school, with only a single impact remaining at the end of 3rd grade for children in each age cohort: a favorable impact for the 4-year-old cohort (ECLS-K Reading) and an unfavorable impact for the 3-year-old cohort (grade promotion). The scores of the Head Start and control group children remained lower than the norm for the population.

For mathematics**,** one impact was found on a single outcome measure, for the 3-year-old cohort only, at one time point (end of their Head Start year).

Social-Emotional Development. With regard to children's social-emotional development, the results differed by age cohort and by the source of the information on the

child's behavior. For children in the 4-year-old cohort, there were no observed impacts through the end of kindergarten and then favorable impacts reported by parents and unfavorable impacts reported by teachers at the end of 1st and 3rd grades. Parents reported less withdrawn behavior at the end of 1st grade and less aggressive behavior and less total behavior problems at the end of 3rd grade. In contrast, at the end of 1st grade, teachers reported more shy behavior and more problems in their interactions with the Head Start children. At the end of 3rd grade, teachers reported more problems in their relationship with Head Start children and a lower percentage of Head Start children in the normal category for emotional symptoms. Children's own reports showed one unfavorable impact at the end of 3rd grade (peer relations).

In contrast to the 4-year-old cohort, the favorable social emotional impacts found for the 3-year-old cohort in the early years of the study continued into early elementary school. There were favorable impacts at all data collection points through the end of 3rd grade on several parent-reported measures of children's social-emotional development. Parents reported less hyperactive and total problem behaviors at the end of the age 3 year, better social skills and positive approaches to learning at the end of the age 4 year and kindergarten as well as less hyperactive behavior at the end of kindergarten. At the end of 1st grade, there is evidence of greater parent-child closeness and improved child-parent relationships for Head Start children compared to the control group. Finally at the end of 3rd grade, parents again reported better social skills and positive approaches to learning for Head Start children. However, there were no impacts on teacher-reported measures of social-emotional development for the 3-year-old cohort at any data collection point or on the children's self-reports in 3rd grade.

Health. In the health domain, early favorable impacts were noted for both age cohorts, but by the end of 3rd grade, there were no remaining impacts for either age cohort. There were favorable impacts on children's receipt of dental care for both age cohorts during their preschool years. There also were favorable impacts on the availability of health insurance for children in both age cohorts at the end of kindergarten and for the 4-year-old cohort at the end of 1st grade.

Parenting Practices. Finally, with regard to parenting practices, impacts were concentrated in the younger cohort. For the 4-year-old cohort, there were limited favorable

impacts, with one impact reported at the end of the Head Start year (less time out[80]) and one impact at the end of the 3rd grade (more time spent with child). For the 3-year-old cohort, there was (1) less use of spanking and an increase in the frequency of parents' reading to their child and involving them in cultural enrichment activities at the end of the Head Start year, (2) less use of the authoritarian parenting style (characterized as high control and low warmth) at the end of the age 4 year, (3) less spanking and use of time out at the end of kindergarten, (4) less use of time out and less use of the authoritarian parenting style at the end of 1st grade, and (5) an increase in the use of the authoritative parenting style (characterized as high control and high warmth) at the end of 3rd grade. These impacts on parenting practices are particularly relevant, given the pattern of favorable impacts on social emotional outcomes for the 3-year-old cohort.

Is There Variation in Impacts on Children and Families?

In addition to looking at Head Start's average impact across the diverse set of children and families who participate in the program, this study also examined how impacts varied among different types of participants. There is evidence that for some outcomes, Head Start had a differential impact for some subgroups of children over others. The subgroup findings are exploratory, given the number of statistical tests conducted, but are worth further consideration in follow-up research.

At the end of 3rd grade, the most striking subgroup finding was related to children from high risk households. For this subgroup, children in the 3-year old cohort demonstrated sustained cognitive impacts across all the years from pre-K through 3rd grade. At the end of 3rd grade, the Head Start children from high risk households showed favorable impacts on the ECLS-K Reading Assessment, the WJIII Letter-Word Identification, and the teacher-reported reading/language arts skills. This was in contrast to the impacts for children in lower and moderate risk households, for whom there were no impacts. Those children who started out with more familial stressors than their peers were found to have multiple positive impacts on the direct student assessments over time. Also among the 3-year-old cohort, children of parents with no reported depressive symptoms experienced sustained benefits of Head Start in the cognitive

[80] The interpretation of a decrease in the use of time out as a disciplinary technique is not clear. Parents may have decreased the use of time out because their children's behavior improved, or they may have switched to other (potentially less positive) parenting techniques.

domain through the end of 3rd grade and in the social-emotional and parenting practices domain through the end of 1st grade.

Among the 4-year-olds, the subgroups that demonstrated sustained benefits were children of parents who reported mild depressive symptoms, severe depressive symptoms, and Black children. Head Start children of parents reporting mild depressive symptoms demonstrated favorable cognitive impacts through the end of 3rd grade. This is in contrast to those whose parents reported, no, moderate, or severe depressive symptoms. However, favorable impacts were reported at the end of the Head Start year for parents with severe depressive symptoms. In the social-emotional and the parenting domains, predominantly favorable parent-reported impacts were sustained for children of parents with severe depressive symptoms. Black children experienced favorable impacts in the social-emotional domain at the end of kindergarten through 3rd grade as reported by teachers, parents, and the child self-report.

Finally, several subgroups were identified that experienced solely–or primarily–unfavorable impacts of Head Start that were sustained through 3rd grade. For the 4-year-old cohort, these groups included White children with unfavorable impacts in the social-emotional domain and for the 3-year-old cohort, children of parents with mild depressive symptoms also with unfavorable impacts in the social-emotional domain. Many subgroups in both age cohorts experienced a mixture of favorable and unfavorable impacts, particularly in the social-emotional domain.

Some of the subgroup impacts from earlier years were not sustained through 3rd grade. For example, the favorable social-emotional impacts for children in the 4-year-old cohort of parents with mild depressive symptoms, and the favorable cognitive impacts found at the end of 1st grade for children in the 3-year-old cohort from non-urban settings, were not sustained through 3rd grade.

Discussion

Head Start has the ambitious mandate of improving educational and developmental outcomes for children from economically disadvantaged families. Head Start's mandate requires that it meet the needs of the whole-child, including the cognitive, social-emotional, and health needs of children, and positively influence the parenting practices of their parents. This study

examined the impacts of Head Start on these four domains and whether earlier impacts were sustained into 3rd grade.

The lasting effects of Head Start and early childhood education in general on children's outcomes have been the focus of much study. Considering only outcomes through early elementary school and middle childhood, results for the HSIS cognitive outcomes are in line with other experimental and non-experimental early education studies. Non-experimental Head Start studies showed initial positive impacts of a roughly similar magnitude to those found in the HSIS that dissipated as the children entered early elementary school (Currie & Thomas, 1995; Garces, et al., 2002; Ludwig & Phillips 2008; Deming 2009). Moreover, recent longitudinal data from the experimental evaluation of Early Head Start (Vogel et al., 2010) showed a similar pattern of early positive impacts that were not sustained into elementary school. Experimental results from the HighScope Preschool Curriculum Comparison study found negligible differences between study groups in cognitive and academic outcomes in the first decade of study (Schweinhart & Weikart, 1997). Similar conclusions about the size and lack of persistence of early impacts were reported in a recent broader meta-analysis of early childhood interventions (Leak et al., 2010). However, as we discuss later, some studies, including those that did not show differences in elementary school, reported finding positive effects later in adulthood. Although the underlying cause of the rapid attenuation of early impacts is an area of frequent speculation, we don't have a good understanding of this observed pattern. All we can say is after the initially realized cognitive benefits for the Head Start children, these gains were quickly made up by children in the non-Head Start group.

We do not yet know if there will be positive outcomes for HSIS participants later in life, however, research suggests that positive outcomes later in life are possible. Despite a growing body of research about relatively rapid dissipation of early cognitive impacts, there is some evidence suggesting that positive effects of Head Start may have an impact on participants' later life such as later school success and early adulthood outcomes (Garces, et al., 2002; Ludwig & Miller, 2007; Deming, 2009). Garces, Thomas, and Currie (2000) conducted a non-experimental study that reported evidence of long-term improvement for Head Start participants on outcomes such as school attainment, earnings, and crime reduction, for some race and gender combinations. Ludwig and Miller (2007), using a regression discontinuity design, reported that increases in Head Start funding were associated with a decline in mortality rates for children

ages five to nine from causes of death that could be affected by the program, an increase in high school completion, and an increase in the likelihood of attending some college. Both of these findings were based on Head Start programs that operated in the 1960s through the 1980s. More recently, Gelber and Isen (2011), using the HSIS data, reported that parents of children assigned to Head Start were more involved with them in a variety of activities both during Head Start enrollment and the early elementary years. The authors suggested that increases in parent involvement may mediate long-term impacts on child outcomes. According to a recent paper by Gibbs, Ludwig, & Miller (2011) such delayed or "sleeper" effects may occur because of the Head Start benefits in the area of children's social and emotional development, i.e., improved socialization and emotional strength may have later school-related payoffs.

Research from non-Head Start samples with similar populations also suggests that "sleeper" effects may present years after exposure to early education. Using data from the randomized study, Project Star (1985-89 Tennessee K-3 Class Size Study), Chetty et al. (2010) reported that improvements in kindergarten test scores translated into higher lifetime earnings, more likely college attendance, retirement savings, home ownership, and residence in a better neighborhood. Children from the HighScope group completed more years of school, had less self-reported misconduct at age 15, fewer felony arrests, and fewer property crime arrests than those who received direct instruction. Initially, no early academic differences were found, but the long-term impacts suggest benefits of quality early childhood education in early adulthood outcomes. Although Project Star and the HighScope Preschool Curriculum Comparison study were not focused on Head Start, like the other Head Start studies, they point to the importance of early education for improving children's long-term outcomes.

In addition to considering the possible long-term impacts, there are a few other things to consider in interpreting the findings of this study. First, this was not a comparison of Head Start to parental care. This study evaluated the Head Start program against a mixture of alternative care settings rather than against a "no services" condition. About 40 percent of the control group did not receive formal preschool education, and, for those who did, quality was generally lower than in Head Start. Nevertheless, many of the control group children received services. Further, among those who participated in non-parental care, the control group children were actually in non-parental care for more hours than the Head Start group—on average, children in the control group attended some type of non-parental care about four to five hours

more per week in the Head Start year, compared with children who had access to Head Start. Consequently, to achieve measurable impacts, Head Start (as noted above) had to outperform services that children in the control group received.

Additionally, to date the findings do not differentiate impacts for children who received differing quality in Head Start. Although the quality is high on average, Head Start programs varied in terms of academic instruction in the key areas measured as part of this study, i.e., early development of language and literacy and mathematics skills. This is not to say that all Head Start programs were not trying their best to improve children's development in these areas, but rather on average the program may not have been potent enough in this particular domain to provide the level of overall learning gains needed to move children into a different, and more rapid, growth trajectory. The pattern for the HSIS data showed initial accelerated gains for the Head Start children, then these gains were quickly made up by the control group children, followed by continued gains at the same pace for both groups. The variation in quality may have contributed to the lack of statistically significant differences in the cognitive domain in the early elementary grades.

A separate report will explore how variation in Head Start quality is related to children's impacts as well as how children's later experiences in the school and community affect their outcomes at 1st and 3rd grade, including whether some later experiences help to sustain impacts through the early elementary grades.

Finally, this study leaves many important questions about Head Start unanswered. These questions include, but are certainly not limited to: Is there a benefit to having two years of Head Start rather than one year? What accounts for the subgroup patterns observed in this report? The Head Start Impact Study is an excellent database for methodological and child development research due to its size, longitudinal data, and multiple variables. Hopefully, researchers will take advantage of the data from this study, which will be made available through a data archive,[81] to further the understanding of the role Head Start plays in the well-being of children and families.

[81] The 3rd grade data will be archived at the Child Care Early Education Research Connections Project. http://www.childcarereserach.org/childcare/welcome.

References

Achenbach, T.M., Edelbrock, C., & Howell, C.T. (1987). Empirically based assessment of the behavioral/emotional problems of 2-3-year old children. *Journal of Abnormal Child Psychology, 15,* 629-650.

Barnett, W.S., Hustedt, J.T., Robin, K.B., & Schulman, K.L. (2004). *The state of preschool. 2004 state preschool yearbook.* New Brunswick, NJ: The National Institute for Early Education Research.

Barnett, W.S., Carolan, M.E., Fitzgerald, J., & Squires, J.J. (2011). *The state of preschool 2011. State preschool yearbook.* New Brunswick, NJ: The National Institute for Early Education Research.

Baumrind, D. (1971). Cultural patterns of paternal authority. *Developmental Psychology Monograph, 4*.

Benjamini, Y. & Hochberg, Y. (1995). Controlling for the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological), 57*(1), 289-300.

Block, J. H. (1965). *The child-rearing practices report (CRPR): A set of Q items for the description of parental socialization attitudes and values.* Berkley, CA: University of California at Berkley, Institute of Human Development.

Bloom, H. S. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Review, 8*(2), 225-246.

Cappella, E. & Larner, M. B. (1999). America's schoolchildren: Past, present and future. *The Future of Children, 9*(2), 21-29.

Chetty, R., Friedman, J.N., Hilger, N., Saez, E., Schanzenbach, D.W., & Yagan, D. (November 2010). $320,000 kindergarten teachers. *Kappan,* 92(3), 22-25.

Cunha, F., Heckman, J.J., & Schennach, S.M. (2006). *Estimating the technology of cognitive and noncognitive skill formation.* Unpublished manuscript, University of Chicago, Department of Economics. Presented at the Yale Conference on Macro and Labor Economics, May 5-7, 2006.

Currie, J. & Thomas, D. (1995). Does Head Start make a difference? *American Economic Review,* 85(3), 341-364.

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from Head Start. *American Economic Journal of Applied Economics,* 1(3), 111-134.

Duncan, G.J., Dowsett, C.J., Claessens, A., Magnuson, K., Huston, A.C., Klebanov, P., Pagani, L.S., Feinstein, L, Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, 43(6), 1428-1446.

Dunn, L.M., Dunn, L.L., & Dunn, D.M. (1997). *Peabody picture and vocabulary test (PPVT),*, third edition. Circle Pines, MN: American Guidance Services.

Dunn, L.M., Padilla, E.R., Lugo, D.E., & Dunn, L.L. (1986). *Test de vocabulario en imágenes peabody (TVIP)*. Circle Pines, MN: American Guidance Service.

Fast Track Project. (1990). *Social Competency Scale*—(teacher version). University Park, PA: Pennsylvania State University.

Garces, E., Thomas, D., & Currie, J. (December, 2000). *Favorable long-term effects of Head Start.* National Bureau of Economic Research (NBER) Working Paper No. 8054. Cambridge, MA: NBER.

Garces, E., Thomas, D., & Currie, J. (2002). Longer term effects of Head Start. *American Economic Review* 92: 999–1012.

Gelber, A. & Isen, A. (December 2011). *Children's schooling and parent's investment in children: Evidence from the Head Start Impact Study.* NBER Working Paper No. 17704. Cambridge, MA: National Bureau of Economic Research, Inc.

Gennetian, L.A., Morris, P.A., Bos, J.M., & Bloom, H.S. (2005). Constructing instrumental variables from experimental data to explore how treatments produce effects. In H.S. Bloom (Ed.), *Learning more from social experiments: Evolving analytic approaches*. New York: Russell Sage Foundation.

Gibbs, C., Ludwig, J., & Miller, D. (2011). *Does Head Start do any lasting good?* Paper presented at the American Economic Association meeting, January 10, 2011, Denver, CO

Goal One Technical Planning Group. (1991). The Goal One Technical Planning Subgroup report on school readiness. In National Education Goals Panel (Ed.), *Potential strategies for long-term indicator development: Reports of the technical planning subgroups* (Report No. 91-0, pp. 1-18). Washington, DC: National Education Goals Panel.

Goal One Technical Planning Group. (1993). *Reconsidering children's early development and learning: Toward shared beliefs and vocabulary. Draft report to the National Education Goals Panel.* Washington, DC: National Education Goals Panel.

Goodman R. (1997) The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38, 581-586.

Heckman, J.J. (2008). School, skill, and synapses. *Economic Inquiry*, 46(3), 289-324.

Kagan, S.L., Moore, E., & Bredekamp, S. (Eds.). (1995). *Considering children's early development and learning: Toward shared belief and vocabulary.* Washington, DC: National Education Goals Panel.

Johnson, H.P., O'Brien-Strain, M., & Acumen LLC. (2000). *Getting to know the future customers of the Office of Child Support: Projections report for 2004 and 2009.* Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families.

Leak, J., Duncan, G.J., Li, W., Magnuson, K., Schindler, H., & Yoshikawa, H. (2010). *Is timing everything? How early childhood education program impacts vary by starting age, program duration and time since the end of the program.* Paper presented at the Association for Policy Analysis and Management meeting, November 4-6, 2010, Boston, MA.

Ludwig, J. & Miller, D.L. (2007). Does Head Start improve children's life chances? Evidence from a regression discontinuity design. *Quarterly Journal of Economics*, 122: 159–208.

Ludwig, J. & Phillips, D.A. (2008). Long-term effects of Head Start on low-income children. *Annals of the New York Academy of Sciences*, 1136: 257-268.

Marsh, H. (1990). *Self-description questionnaire manual*. Campbelltown, N.S.W., Australia: University of Western Sydney, Macarthur.

Najarian, M. & Yan, F. (November, 2008). *Summary of the Head Start Impact Study spring 2008 third grade data analysis*. Princeton, NJ: Educational Testing Service.

Pianta, R.C. (1992). *Child-parent relationship scale*. Charlottesville, VA: University of Virginia.

Pianta, R.C. (1996). *Student-teacher relationship scale*. Charlottesville, VA: University of Virginia.

Pollack, J., Atkins-Burnett, S., Rock, D., and Weiss, M. (2005). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K), Psychometric report for the third grade* (NCES 2005–062). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Schweinhart, L.J. & Weikart, D.P. (1997). The HighScope preschool curriculum comparison study through age 23. *Early Childhood Research Quarterly, 12,* 117–143.

U.S. Department of Health and Human Services, Administration for Children, Youth, and Families. (January, 1999). *Evaluating Head Start: A recommended framework for studying the impact of the Head Start program.* Washington, DC: Author.

U.S. Department of Health and Human Services, Administration for Children, Youth, and Families. (2001). Head *Start program performance standards and other regulations*. Washington, DC: Author.

U.S. Department of Health and Human Services, Administration for Children and Families. (2004). Head Start Program Fact Sheet. Retrieved September 14, 2009 from the U.S. Department of Health and Human Services Website: http://www.acf.hhs.gov/programs/ohs/about/index.html

U.S. Department of Health and Human Services. Administration for Children and Families. (January 2010). *Head Start Impact Study. Final Report.* Washington, DC: Author.

U.S. Department of Health and Human Services. Administration for Children and Families. (January 2010). *Head Start Impact Study. Technical Report.* Washington, DC: Author.

U.S. General Accounting Office. (1997). *Head Start: Research provides little information on impact of current program*. Report to the Chairman, Committee on the Budget, House of Representatives. Washington, DC: Author.

U.S. General Accounting Office. (1998*). Head Start: Challenges in monitoring program quality and demonstrating results*. Report to Congressional Requestors. Washington, DC: Author.

Ventura, S.J. (2009). *Changing patterns of nonmarital childbearing in the United States*. National Center for Health Statistics Data Brief. Atlanta, GA: Centers for Disease Control and Prevention.

Vogel, C.A., Xue, Y., Moiduddin, E., Kisker, E., & Carlson, B.L. (2010). *Early Head Start children in grade 5: Long-term follow-up of the Early Head Start research and evaluation project study sample.* OPRE Report #2011-8, Washington DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office for Planning, Research, and Evaluation.

Warren, J.R. & Saliba, J. (June, 2012). *First through eighth grade retention rates for all 50 states: A new method and initial results.* Paper presented at the Annual Meeting of the Population Association of America in San Francisco, CA.

Woodcock, R.W. & Muñoz-Sandoval, A.F. (1996). *Bateriá Woodcock-Muñoz pruebas de aprovechamiento-Revisada*. Itasca, IL: Riverside Publishing.

Woodcock, R.W., McGrew, K.S., & Mather, N. (2001). *Woodcock-Johnson III tests of achievement*. Itasca, IL: Riverside Publishing.

Appendix A

Updated Weighting and Analysis Procedures

The 3rd Grade Follow-up to the Head Start Impact Study implemented the same data collection, weighting, and analysis procedures used for the Head Start Impact Study with few exceptions. Any changes to the weighting and analysis procedures for the 3rd grade follow-up report are documented in this Appendix. For details of the data collection, weighting, and analysis procedures used for the Head Start Impact Study, please see the *Head Start Impact Study Technical Report* (2010).[1]

Updated Weights for the 3rd Grade Follow-up to the Head Start Impact Study

Weights were calculated for each of the children that were sampled in 2002 and randomly assigned to treatment and control groups. Since assignment to treatment and control groups was random from the Head Start applicant lists, theoretically the only difference between the two groups is that the Head Start group was allowed access to attend Head Start in Fall 2002, while the control group was not.

Baseline data collection began in Fall 2002, and the treatment and control groups were assessed every spring thereafter through the 3rd grade (see Exhibit A.1). Cross-sectional weights were calculated for Fall 2002 and for each spring through 3rd grade (see Exhibit A.2). When properly weighted, each of the treatment and control groups represents the national population of children who were new applicants to Head Start in 2002 and are in 3rd grade in spring 2007 (age 4 cohort) or 2008 (age 3 cohort).

The weights are intended to be used with data collected from the child assessment (CA), the parent interview (PI), the teacher survey/teacher child rating (TS/TCR), and the principal survey (PS) in the 3rd grade follow-up. Six weights were produced for each child: a child assessment weight, parent interview weight, teacher survey/TCR reading weight, teacher survey/TCR math weight, teacher survey socio-emotional outcomes ("all-purpose weight"), and the principal survey for the school the child attended. For the "all-purpose weight" the children from spring 2007 and 2008 were pooled instead of producing a separate weight for each year.

---

[1] U.S. Department of Health and Human Services, Administration for Children and Families (January 2010). *Head Start Impact Study: Technical Report. Washington, DC.* http://www.acf.hhs.gov/programs/opre/hs/impact_study/reports/impact_study/hs_impact_study_tech_rpt.pdf.

Note that the unit of analysis is always the child, regardless of the instrument used to collect the data.

The final child weight for each instrument reflects the probability of selection at each stage of sampling and can be written as the product of these component weights:

> Final Child Weight = PSU weight x HS Program weight x Center weight x Within-Center Child Base Weight x Child Nonresponse Adjustment Factor x Poststratification Factor x Trimming Factor.

The calculation of the weights is discussed in the sections below.

Exhibit A.1. Head Start Impact Study Data Collection

Age Cohort	Fall 2002	Spring 2003	Spring 2004	Spring 2005	Spring 2006	Spring 2007	Spring 2008
3	HS	HS	HS	K	1st grade	No data collected	3rd grade
4	HS	HS	K	1st grade	No data collected	3rd grade	No data collected

Key: HS indicates Head Start and K indicates kindergarten

Exhibit A.2. Cross-sectional Weights Produced for 3rd Grade Head Start Follow-up Analysis

Weight	Definition of Respondent Child	Spring 2007	Spring 2008
Child Assessment	Completed Child Assessment	X	X
Parent Interview	Completed Parent Interview	X	X
Teacher Survey/Teacher Child Rating Reading Assessment	Completed TS and TCR reading assessment	X	X
Teacher Survey/Teacher Child Rating Math Assessment	Completed TS and TCR math assessment	X	X
Principal Survey	Completed Principal Survey	X	X
Teacher Survey All-Purpose Weight (Socio-Emotional Outcomes)	Completed reading or math assessment, or both.		X

Base Weights

A base weight was calculated for each sampled child based on their overall probability of selection, including the sampling of geographic areas (PSUs), Head Start grantees/delegate agencies (programs), and centers. The overall child base weight can be written as the product of the weight at each stage of sampling:

Overall Child Base Weight = PSU weight x program weight x center weight x (within-center child weight) where the within-center child 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 (treatment) group children, and as

$$\frac{\#\text{ newly Entering Age 3 \& 4 Enrollment in Center}}{\#\text{ control children sampled in center}}$$

for the control group children. The newly entering enrollment in the numerator of the base weight was obtained from the centers for Fall 2002 or from the previous year if not available. The calculation of the PSU, program and center weights is discussed in detail in Chapter 2 (Analytical Sampling Weights) in the *Head Start Impact Study Technical Report* (2010).[2]

Nonresponse Adjustments

The base weights were adjusted for nonresponse to the 3rd Grade Follow-up child assessment, parent interview, principal survey, and teacher survey. The teacher survey weights were adjusted separately for the math, reading and socio-emotional evaluation portions of the teacher survey. This resulted in six weights for each child.

To capture the variation in response rates, cells were formed based on child and Head Start program characteristics that are correlated with response rates. The tree structure identified by CHAID (Chi-squared Automatic Interaction Detector) software was used to identify these variables and create the nonresponse adjustment cells. The variables identified as correlates with nonresponse by CHAID are given for each instrument in Exhibit A.3. A separate nonresponse adjustment factor was calculated within each cell. The nonresponse adjustment factor is equal to the inverse of the weighted response rate for the cell, using the child base weights. The adjustment factor helps control for nonresponse bias by compensating for different

[2] See Footnote 1.

Exhibit A.3. Variables Used to Form Nonresponse Adjustment Cells within Treatment, Control Groups

Weight	No Show (Y/N)	Cross-over (Y/N)	PSU Group-ing	Child's Lan-guage (English vs. Other)	Child's Gender	Child's Race	Mother's Educa-tion	Head Start Program (FT/ PT)	Head Start Program Metro Status	State	Head Start Region	Head Start-like Programs in State
Child Assessment												
2007	X		X	X			X	X	X			
2008	X		X	X								
Parent Interview												
2007	X		X	X			X		X	X		
2008	X	X	X		X					X		
Teacher Reading Assessment												
2007	X		X				X			X	X	X
2008	X	X	X	X		X	X					
Teacher Math Assessment												
2007	X		X		X		X		X	X		
2008	X	X	X	X	X				X			
Principal Survey												
2007	X		X				X		X	X		
2008	X		X	X		X		X				
Social Emotional	X	X	X	X	X				X		X	

Exhibit A.4. Response Rates for 3rd Grade Follow-up to the Head Start Impact Study

Weight	Number of Respondents		Unweighted Response Rate		Weighted Response Rate	
Child Assessment	Treatment	Control	Treatment	Control	Treatment	Control
2007	917	594	73	69	72	71
2008	1,183	717	77	70	79	71
Parent Interview						
2007	978	623	78	73	77	75
2008	1,240	753	81	73	83	75
Teacher Reading Assessment						
2007	713	454	57	53	58	57
2008	933	560	61	55	63	55
Teacher Math Assessment						
2007	732	460	59	54	59	57
2008	934	578	61	56	62	57
Principal Survey						
2007	815	511	65	60	66	63
2008	965	590	63	58	66	58
Social Emotional Outcomes	1,742	1,090	63	58	63	60

data collection response rates across various demographic and geographic groups of children. It spreads the weight of the nonresponding children over the responding children in the cell, so that they represent the children who did not complete the instrument as well as children who weren't sampled. This maintains the same mix of the sample across cells as would have been present had there been no nonresponse.

Poststratification

To reduce the sampling error for estimates of the newly entering Head Start population, the nonresponse-adjusted weights of children in the age 4 cohort were poststratified to the fall 2003 Head Start National Reporting System (HSNRS) newly entering enrollment totals for 4-year-olds by race/ethnicity (comparable totals for 3-year –olds were not available). The three race/ethnicity categories were Hispanic, non-Hispanic/Black, and White/Other. The adjustment factor was calculated as the ratio of the proportion of HSNRS total newly entering age 4 enrollment in the race/ethnicity category to the sample estimate of this proportion using the 84 sampled Head Start programs with their first year age 4 enrollment as reported in the HSNRS. The poststratification factors were .80 for Hispanic, 1.45 for Black, and 1.036 for White/other,

indicating an overrepresentation of age 4 Hispanic children and underrepresentation of Black children in the Head Start Impact sample as compared to the HSNRS.

Trimming

A final trimming adjustment was made for extremely large child weights. These outlier weights were trimmed back to four times the average weight to avoid large sampling errors. This introduces a very small amount of bias into the survey estimates, in exchange for a reduction in the total mean square error (MSE). The percent of respondents that had their final weights trimmed ranged from 1.1 to 2.5 percent, depending on the instrument and year (spring 2007 or spring 2008).

Variance Estimation

Estimates from the 3rd Grade Follow-up 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 children who applied for their first year of Head Start in Fall 2002. This type of error is known as sampling error or variance. 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, which all must be reflected in the standard error calculations.

A set of 76 jackknife replicate weights was created for each child corresponding to each of the six full-sample weights, for use in the calculation of standard errors. The replicate weights can be used with software designed for the analysis of complex survey data, such as SUDAAN, WesVar, Stata, and the survey procs in SAS (e.g. proc surveyreg). For multi-level modeling with weights, the HLM software package is available (version 6 or later) and Stata version 12. The Taylor Series linearization method can also be used with any of the software packages above to calculate standard errors using the variance strata and variance pseudo-PSU

codes placed on the analysis files. SPSS Complex Samples can be used with the linearization method as well.

Updated Analysis Procedures for the 3rd Grade Follow-up to the Head Start Impact Study

Defining and Balancing the Analysis Sample

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 teachers are weighted and analyzed according to the children they describe. This makes all impact findings representative of all newly entering Head Start children in the nation in 2002 in communities in which there were more potential program participants than funded Federal Head Start slots and in programs that meet the study criteria for eligibility.

The annual cross-sectional samples are chosen to maximize the data available for analysis each spring. Thus, they include every completed child assessment, parent interview, or teacher-child report for that year. In each instance, the comparability of the treatment and control group samples established at random assignment is maintained to the greatest extent possible by adjusting the initial sampling weights to offset observable differences between respondents and non-respondents for each cohort.

A comparison of the characteristics of the children and families in the 3rd grade impact analysis sample, using characteristics measured at baseline in fall 2002, is presented in Appendix B for both age cohorts. The comparisons use baseline data on demographic characteristics and baseline data on variables measured as outcomes in later years (i.e., pretest measures). Tables are organized to present the comparisons by the respondent source, i.e., for the direct child assessment outcomes, for parent-reported outcomes, and for outcomes reported by teachers and principals. In these tables, observations are weighted to reflect the share of the national population they represent, i.e., the 3rd grade impact analysis sample seeks to replicate this population as accurately as possible even though not every child can be included. Any differences between the program and control groups in these tables reflect the combined consequences of (1) chance differences between the treatment and control groups created at random assignment, (2) differential nonresponse in the 3rd grade data collection following

weighting adjustment, and (3) for pretest measures, the possible early impacts of Head Start in fall 2002 before baseline data could be collected.

Adjusting for differences between the treatment and control group analysis samples by including the fall 2002 measures in the 3rd grade impact regressions will offset two of these factors. In particular it will reduce the threat of nonresponse bias and increase the statistical precision of the impact findings by offsetting chance differences present at baseline that continue to influence 3rd grade outcomes. Unfortunately, it will also cause the 3rd grade estimates to omit the earliest impacts of Head Start on participating children if any of those occurred prior to collection of included fall 2002 measures. Most of the fall 2002 data were collected during a three-month period from October 2002 through December 2002 (with most collected by mid-November) at a lag from the start of Head Start participation in the treatment group in August and September 2002. If impacts of Head Start occurred quickly that fall, inclusion of the fall measures as covariates in the 3rd grade impact regressions will attenuate the impact estimate—i.e., bias it toward zero—since the portion of the impact achieved prior to fall 2002 data collection would be removed from the 3rd grade findings.

In our judgment, attenuation of the 3rd grade impact findings through removing early impacts of Head Start from estimates by including pre-test variables would do more harm than excluding pre-test variables, thereby failing to adjust for chance differences at random assignment or any nonresponse differences that remain in the outcome data following reweighting.[3] Reflective of this perspective, we include pretest measures as covariates in the impact regressions only if strong evidence exists that an early impact of Head Start *did not occur* prior to collection of the candidate pretest in fall 2002. To seek that evidence, we estimated impacts on the candidate fall 2002 measures in standard deviation units (i.e., as effect sizes) and formed 90-percent confidence intervals around them. Where the 90-percent confidence interval fit entirely between -0.20 and 0.20 standard deviations—i.e., where we could be 90-percent certain any early impact of Head Start was small—we deemed the pretest measure appropriate

[3] Holding off from adjusting for chance differences between the treatment and control groups at baseline affects only the statistical precision, not the biasedness or unbiasedness of the 3rd grade impact estimates. Failing to make adjustment for differences created by nonresponse in follow-up data collection does affect bias, but not in a known direction. In contrast, attenuation due to removal of early impacts through inclusion of baseline covariates will consistently bias impact estimates toward zero, consistently understating the impact of the Head Start program.

for inclusion as a covariate in 3rd grade impact regressions. Exhibit A.5 lists the fall 2002 pretest measures that met this condition for each of the study's age cohorts.

Exhibit A.5. Fall 2002 Pretest Measures Found Appropriate for Inclusion as Covariates in 3rd Grade Impact Regressions, by Age Cohort

4-Year-Old Cohort	3-Year-Old Cohort
Peabody Picture Vocabulary Test (PPVT)	Peabody Picture Vocabulary Test (PPVT)
NA	Color Naming
Counting Bears	Counting Bears
NA	McCarthy Draw-a-Design
Social Skills and Positive Approaches to Learning	Social Skills and Positive Approaches to Learning*
NA	Total Behavior Problems
NA	Aggressive Behavior
Withdrawn Behavior	Withdrawn Behavior
NA	Social Competencies
Parent Used Time Out	NA
Parent Spanked Child	NA
Parental Safety Practices	NA
Safety Practices	Safety Practices
Family Cultural Enrichment	Family Cultural Enrichment*
Child Health Status is Excellent or Very Good	Child Health Status is Excellent or Very Good
NA	Child Has Health Insurance*
NA	Child Had Injury Care in Last Month

An * indicates the measure was used as a covariate in all 3-year-old cohort impact regressions.

Not all of these pretest measures are used as covariates in every 3rd grade impact regression. For the 3-year-old cohort, three such measures were included in all regressions because they differed significantly (at the .05 significance level) between the treatment and control group portions of the 3rd grade analysis sample, as designated by an * in Exhibit 2.8: social skills and positive approaches to learning, family cultural enrichment, and child has health insurance. It is particularly important to "equalize" these factors when computing impacts since they represent treatment-control asymmetries in the baseline characteristics of the outcome samples that nonresponse weighting did not remove. In addition to these universal inclusions for the 3-year-old cohort, selected pretest measures were added to the impact regression for specific 3rd grade outcome variables in each age cohort. These pretest measures, shown in Exhibit A.6,

were thought to be particularly helpful in predicting the associated outcome variable and hence in reducing nonresponse bias and increasing statistical precision. Note that for some outcomes no pretest measure was included since none of the measures in Exhibit A.5 seemed likely to predict that particular outcome.

Exhibit A.6. Pretest Measures Used in 3rd Grade Impact Analyses, by Outcome Measure

Outcome	Baseline Covariate
COGNITIVE	
ECLS-K Reading	PPVT
PPVT (Peabody Picture Vocabulary Test-adapted)	PPVT
WJIII Letter Word Identification	PPVT
WM Letter Word Identification	None
WJ III Applied Problems	Counting Bears
WJIII Calculation	Counting Bears
Promotion	PPVT
Language and Literacy Ability	PPVT
Math Ability	Counting Bears
Reading/Language Arts Skills	PPVT
Math Skills	Counting Bears
SOCIAL-EMOTIONAL	
Aggressive Behavior	Social Skills and Positive Approaches To Learning
Hyperactive Behavior	Social Skills and Positive Approaches To Learning
Withdrawn Behavior	Social Skills and Positive Approaches To Learning
Total Behavior Problems	Social Skills and Positive Approaches To Learning
Social Skills and Positive Approaches To Learning	Social Skills and Positive Approaches To Learning
Conduct Problems-Percent in Normal Category	Social Skills and Positive Approaches To Learning
Emotional Symptoms-Percent in Normal Category	Social Skills and Positive Approaches To Learning
Hyperactivity-Percent in Normal Category	Social Skills and Positive Approaches To Learning
Peer Problems-Percent in Normal Category	Social Skills and Positive Approaches To Learning
Pro-social Behavior-Percent in Normal Category	Social Skills and Positive Approaches To Learning
Total Difficulties-Percent in Normal Category	Social Skills and Positive Approaches To Learning
Closeness with Teacher	Social Skills and Positive Approaches To Learning
Conflict with Teacher	Social Skills and Positive Approaches To Learning
Positive Teacher-Child Relationships	Social Skills and Positive Approaches To Learning
Social Competency	Social Skills and Positive Approaches To Learning
Externalizing	Social Skills and Positive Approaches To Learning

Exhibit A.6. Pretest Measures Used in 3rd Grade Impact Analyses, by Outcome Measure (continued)

Outcome	Baseline Covariate
Internalizing	Social Skills and Positive Approaches To Learning
Peer Relations	Social Skills and Positive Approaches To Learning
School	Social Skills and Positive Approaches To Learning
HEALTH	
Child Received Dental Care	Child's Overall Health Status Is Excellent/Good
Child Has Health Insurance Coverage	Child's Overall Health Status Is Excellent/Good
Child's Overall Health Status Is Excellent/Good	Child's Overall Health Status Is Excellent/Good
Child Needs Ongoing Care	Child's Overall Health Status Is Excellent/Good
Child Had Care for Injury Last Month	Child Had Care for Injury Last Month
PARENTING PRACTICES	
Parent Spanked Child in Last Week	None
Parenting Style: Authoritarian	None
Parenting Style: Authoritative	None
Parenting Style: Neglectful	None
Parenting Style: Permissive	None
Supportive School Environment	PPVT
Doing Things Together	None
Time Spent with Child	None
Parent Perception of School Services	None
Effect of Parenting on Parent's Life	None
School Contact and Communication	PPVT
Parent Participation	PPVT

For a detailed discussion of the analysis procedures, please see Chapter 5 (Impact Analysis Methods) in the *Head Start Impact Study Technical Report.*[4] The chapter provides detailed information on covariates, imputations, annual cross-sectional impact estimation methods both intent to treat (ITT) and impact on the treated (IOT) estimates, subgroup analyses, and the treatment of no shows and crossovers in the analysis.

[4] See Footnote 1.

Appendix B

Baseline Characteristics for the 3rd Grade Analysis Sample

Exhibit B.1A. Covariate Comparison of 3rd Grade Sample: 4-Year-Old Cohort (Child Assessment)

Variable	Treatment (T)	Control (C)	Difference (T-C)	t-test	p-value	Standard Deviation	Effect Size
Race: Black	0.2599039	0.2443919	0.015512	0.6454814	0.5215088	0.434332	0.0357146
Caregiver Age as of 9/1/2002	29.427234	29.072464	0.3547707	0.8079412	0.422877	7.048114	0.0503355
Family Cultural Enrichment Scale	3.4133177	3.4785422	-0.065225	-0.63675	0.5271353	1.348332	-0.048374
Parent Read to Child in Last Week	0.3859973	0.3584826	0.0275147	0.8230981	0.4142845	0.483473	0.0569106
Parental Safety Practices Scale	3.6648018	3.6505759	0.0142259	0.6037444	0.5486897	0.353756	0.0402138
Parent Spanked Child in Last Week	0.4115242	0.4515947	-0.040071	-1.506711	0.1380543	0.49524	-0.080911
Parent Used Time Out in Last Week	0.6118366	0.6068499	0.0049867	0.1290435	0.8978308	0.487887	0.010221
Both Biological Parents Live with Child	0.5022917	0.5577316	-0.05544	-1.565726	0.1235972	0.499131	-0.111073
Child Age at Time of Spring Assessment	468.86168	469.45901	-0.597326	-0.425159	0.6725091	19.24121	-0.031044
Number of Weeks Between 9/1/2002 and Date of Testing	241.43197	241.43406	-0.002081	-0.010868	0.9913708	3.139659	-0.000663
Child Gender	0.4997962	0.4902776	0.0095186	0.2786703	0.7816247	0.499976	0.0190382
Primary Language Spoken at Home	0.6668547	0.6552695	0.0115852	0.314054	0.7547611	0.47331	0.024477
Biological Mother is a Recent Immigrant	0.2125045	0.2360382	-0.023534	-0.743488	0.4605977	0.41695	-0.056442
Aggressive Behavior	2.8327352	3.1464413	-0.313706	-2.720109	0.0089015	1.703946	-0.184106
Child Received Dental Care	0.8078537	0.6952523	0.1126014	2.8248483	0.0067348	0.431472	0.2609705
Child's Overall Health Status is Excellent/Good	0.7956971	0.7901309	0.0055662	0.1954923	0.8457844	0.405179	0.0137376
Child Had Care for Injury Last Month	0.0572778	0.057728	-0.00045	-0.039287	0.9688146	0.232793	-0.001934
Child Has Health Insurance Coverage	0.8566995	0.9026364	-0.045937	-1.793985	0.0787446	0.325871	-0.140967
Child Needs Ongoing Care	0.1456565	0.2021119	-0.056455	-2.22355	0.0306353	0.378536	-0.149141
Hyperactive Behavior	1.8148251	2.089918	-0.275093	-3.040455	0.0037234	1.538608	-0.178793
Social Competencies	10.835105	10.856146	-0.021041	-0.174705	0.8620035	1.421033	-0.014807
Social Skills and Positive Approaches to Learning	12.405899	12.290887	0.1150112	0.9240804	0.3597962	1.704638	0.0674696
Total Child Behavior Problems	5.9245295	6.5372243	-0.612695	-2.668667	0.0101856	3.723451	-0.16455
Number of Weeks Between 9/1/2002 and Parent Interview	.	.	.	.	.	.	.

Exhibit B.1A. Covariate Comparison of 3rd Grade Sample: 4-Year-Old Cohort (Child Assessment) (continued)

Variable	Treatment (T)	Control (C)	Difference (T-C)	t-test	p-value	Standard Deviation	Effect Size
Withdrawn Behavior	0.716746	0.7876114	-0.070865	-1.144494	0.2577618	1.000216	-0.07085
Mother Gave Birth to Study Child as a Teen	0.1751776	0.1553073	0.0198703	0.9689615	0.3371373	0.371574	0.053476
Race: Hispanic	0.4041033	0.4382902	-0.034187	-0.831204	0.409733	0.493698	-0.069246
Peabody Picture Vocabulary Test (PPVT)	270.12829	272.10195	-1.973662	-0.84291	0.4032149	39.55222	-0.0499
Test de Vocabulario en Imágenes Peabody (TVIP)	264.07099	273.41912	-9.348128	-2.136928	0.0374219	42.00049	-0.222572
Child's Primary Language at Baseline	0.7121037	0.6636571	0.0484466	1.4278788	0.159422	0.463167	0.1045987
Mother Separated/Divorced/Widowed	0.1828096	0.1567673	0.0260423	0.8569739	0.3954684	0.37567	0.0693222
Mother's Education-High School	0.3051383	0.3266446	-0.021506	-0.691062	0.4926601	0.464787	-0.046271
Mother's Education-Less than High School	0.3984715	0.4233031	-0.024832	-0.916686	0.3636217	0.491951	-0.050476
Mother's Education-More than High School	0.2963902	0.2500523	0.0463379	1.3482134	0.1835474	0.445843	0.1039332
Mother Married	0.4488708	0.4703098	-0.021439	-0.688095	0.494511	0.498347	-0.04302
Mother Not Married	0.3683196	0.3729229	-0.004603	-0.122472	0.9030063	0.482959	-0.009531
WJ III Applied Problems	391.52572	388.30983	3.2158881	2.0643523	0.0440882	23.51594	0.1367535
WJ III Letter-Word Identification	308.68829	305.72151	2.966776	1.4045377	0.1662201	25.41065	0.1167533
WM Problemas Aplicados	396.64138	403.23747	-6.596095	-2.263414	0.0278954	22.38105	-0.294718
WM Identificación de letras y palabras	353.35535	358.99902	-5.643674	-2.24825	0.0289109	13.71975	-0.411354
Race: White	0.3359928	0.3173179	0.0186748	0.4832524	0.6309848	0.469057	0.0398136

Exhibit B.1B. Covariate Comparison of 3rd Grade Sample: 4-Year-Old Cohort (Math Teacher Survey-Teacher Child Report)

Variable	Treatment (T)	Control (C)	Difference (T-C)	t-test	p-value	Standard Deviation	Effect Size
Race: Black	0.269718	0.2319118	0.0383324	1.3585367	0.1802724	0.433752	0.0883739
Caregiver Age as of 9/1/2002	29.09793	29.002205	0.1545024	0.3014834	0.7642718	6.657147	0.0232085
Family Cultural Enrichment Scale	3.4786872	3.5449363	-0.06152	-0.48743	0.6280421	1.370011	-0.044904
Parent Read to Child in Last Week	0.4141478	0.3328461	0.0784749	2.0826292	0.0423188	0.483995	0.1621397
Parental Safety Practices Scale	3.6605801	3.6560896	0.0023636	0.0863541	0.9315231	0.353675	0.006683
Parent Spanked Child in Last Week	0.3976257	0.4333014	-0.032696	-0.907996	0.3681512	0.492726	-0.066357
Parent Used Time Out in Last Week	0.5922949	0.6133674	-0.022875	-0.618967	0.5386934	0.489367	-0.046743
Both Biological Parents Live with Child	0.492314	0.5383613	-0.044187	-0.959983	0.3415932	0.499782	-0.088414
Child Age at Time of Spring Assessment	468.14997	470.0363	-1.852743	-1.001345	0.3213877	18.8662	-0.098204
Number of Weeks Between 9/1/2002 and Date of Testing	241.37704	241.18741	0.1548222	0.7125313	0.4793832	3.038293	0.050957
Child Gender	0.496606	0.4941676	0.0031485	0.0798106	0.9367003	0.499979	0.0062973
Primary Language Spoken at Home	0.6626886	0.662181	-0.000405	-0.012029	0.9904492	0.472877	-0.000856
Biological Mother is a Recent Immigrant	0.2357241	0.2424235	-0.006028	-0.155467	0.8770668	0.426467	-0.014135
Aggressive Behavior	2.9194653	3.0272536	-0.110703	-0.888274	0.3785638	1.753243	-0.063142
Child Received Dental Care	0.8102979	0.6859201	0.1232049	2.8905931	0.0056359	0.433212	0.2843985
Child's Overall Health Status is Excellent/Good	0.8127511	0.8207711	-0.004684	-0.171637	0.8644024	0.386944	-0.012104
Child Had Care for Injury Last Month	0.0585057	0.0440016	0.0145847	1.1359472	0.2612886	0.220881	0.0660298
Child Has Health Insurance Coverage	0.8766508	0.9046228	-0.028156	-0.897704	0.3735618	0.312529	-0.09009
Child Needs Ongoing Care	0.1390401	0.1865184	-0.050649	-1.75662	0.0849854	0.368624	-0.137401
Hyperactive Behavior	1.8953939	1.9841429	-0.087327	-0.852392	0.397982	1.52896	-0.057115
Social Competencies	10.792819	10.916103	-0.116116	-0.982352	0.330564	1.424548	-0.081511
Social Skills and Positive Approaches to Learning	12.296398	12.363138	-0.059309	-0.538851	0.592334	1.653361	-0.035872
Total Child Behavior Problems	6.0657053	6.2734316	-0.210188	-0.773725	0.4426654	3.769675	-0.055758
Number of Weeks Between 9/1/2002 and Parent Interview	.	.	.	.	.	.	.
Withdrawn Behavior	0.7081936	0.766216	-0.064633	-1.056778	0.295593	1.013329	-0.063783

Exhibit B.1B. Covariate Comparison of 3rd Grade Sample: 4-Year-Old Cohort (Math Teacher Survey-Teacher Child Report) (continued)

Variable	Treatment (T)	Control (C)	Difference (T-C)	t-test	p-value	Standard Deviation	Effect Size
Mother Gave Birth to Study Child as a Teen	0.1772835	0.1513585	0.0262651	1.097481	0.277586	0.370858	0.0708225
Race: Hispanic	0.3862499	0.4231307	-0.035643	-0.973384	0.3349567	0.490743	-0.072631
Peabody Picture Vocabulary Test (PPVT)	270.35693	272.7782	-2.48477	-0.899689	0.3725143	40.13063	-0.061917
Test de Vocabulario en Imágenes Peabody (TVIP)	265.36426	276.92638	-11.56212	-3.041062	0.003717	44.03209	-0.262584
Child's Primary Language at Baseline	0.7016779	0.6793387	0.021416	0.656578	0.5144041	0.46217	0.0463381
Mother Separated/Divorced/Widowed	0.2049366	0.1752671	0.0300643	1.0054829	0.3194113	0.392672	0.0765632
Mother's Education-High School	0.2977503	0.3108271	-0.015188	-0.425227	0.6724605	0.460036	-0.033015
Mother's Education-Less than High School	0.381944	0.4264023	-0.043192	-1.50093	0.1395399	0.490623	-0.088035
Mother's Education-More than High School	0.3203057	0.2627707	0.0583801	1.5876063	0.1185566	0.454798	0.128365
Mother Married	0.4404468	0.4789329	-0.036806	-1.003802	0.3202132	0.498334	-0.073859
Mother Not Married	0.3546166	0.3458	0.0067421	0.1833648	0.8552389	0.477069	0.0141324
WJ III Applied Problems	393.6601	388.50499	4.9102597	2.361807	0.0220422	22.63828	0.2169007
WJ III Letter-Word Identification	309.37779	304.82849	4.2756268	1.9942422	0.0514848	25.53319	0.1674537
WM Problemas Aplicados	396.60127	403.47338	-6.872108	-2.02119	0.0485239	23.43273	-0.29327
WM Identificación de letras y palabras	352.68331	359.03164	-6.348332	-2.530245	0.0145282	13.81408	-0.459555
Race: White	0.3440321	0.3449575	-0.002689	-0.080881	0.9358531	0.4752	-0.005659

Exhibit B.1C. Covariate Comparison of 3rd Grade Sample: 4-Year-Old Cohort (Parent Interview)

Variable	Treatment (T)	Control (C)	Difference (T-C)	t-test	p-value	Standard Deviation	Effect Size
Race: Black	0.2636348	0.2456105	0.0180244	0.8201791	0.415931	0.435727	0.0413662
Caregiver Age as of 9/1/2002	29.453505	29.256861	0.1966431	0.4964101	0.621738	7.091572	0.0277291
Family Cultural Enrichment Scale	3.4305868	3.4924682	-0.061881	-0.691951	0.4921065	1.357449	-0.045587
Parent Read to Child in Last Week	0.3710435	0.3477225	0.023321	0.7038814	0.4847082	0.479872	0.0485984
Parental Safety Practices Scale	3.6628685	3.6580433	0.0048252	0.2226197	0.8247207	0.350528	0.0137655
Parent Spanked Child in Last Week	0.4132904	0.4386537	-0.025363	-0.892883	0.3761136	0.49446	-0.051295
Parent Used Time Out in Last Week	0.6192137	0.6060694	0.0131443	0.3719534	0.7114686	0.487123	0.0269835
Both Biological Parents Live with Child	0.511385	0.5613905	-0.050006	-1.492238	0.1417973	0.498702	-0.100271
Child Age at Time of Spring Assessment	468.97857	469.79651	-0.817939	-0.588849	0.5585621	19.20654	-0.042586
Number of Weeks Between 9/1/2002 and Date of Testing	.	.	.	.	.	.	.
Child Gender	0.5009854	0.4853178	0.0156676	0.4886214	0.6272043	0.499955	0.0313381
Primary Language Spoken at Home	0.6792795	0.6478015	0.031478	0.8066613	0.4236075	0.472414	0.0666322
Biological Mother is a Recent Immigrant	0.2075008	0.2447891	-0.037288	-1.097624	0.2775241	0.418146	-0.089175
Aggressive Behavior	2.8215492	3.1114385	-0.289889	-2.514729	0.0151076	1.727244	-0.167833
Child Received Dental Care	0.8129091	0.6917734	0.1211357	3.2545361	0.0020196	0.431105	0.2809889
Child's Overall Health Status is Excellent/Good	0.7964351	0.7921918	0.0042433	0.1394259	0.8896628	0.404179	0.0104986
Child Had Care for Injury Last Month	0.0598395	0.0606967	-0.000857	-0.061551	0.9511611	0.237971	-0.003602
Child Has Health Insurance Coverage	0.8611929	0.9022265	-0.041034	-1.586397	0.1188308	0.323325	-0.126911
Child Needs Ongoing Care	0.1432237	0.1997742	-0.056551	-2.342751	0.0230816	0.376564	-0.150175
Hyperactive Behavior	1.8024777	2.0956337	-0.293156	-3.534124	0.0008798	1.527321	-0.191941
Social Competencies	10.830264	10.853215	-0.022951	-0.220592	0.8262907	1.428207	-0.01607
Social Skills and Positive Approaches to Learning	12.390972	12.326403	0.0645693	0.5431703	0.5893789	1.707831	0.0378078
Total Child Behavior Problems	5.8970414	6.4875667	-0.590525	-2.539282	0.0142001	3.71849	-0.158808
Number of Weeks Between 9/1/2002 and Parent Interview	241.49444	241.54491	-0.050475	-0.266229	0.7911362	3.158606	-0.01598
Withdrawn Behavior	0.7071155	0.7679993	-0.060884	-0.991331	0.3262042	0.994326	-0.061231

Exhibit B.1C. Covariate Comparison of 3rd Grade Sample: 4-Year-Old Cohort (Parent Interview) (continued)

Variable	Treatment (T)	Control (C)	Difference (T-C)	t-test	p-value	Standard Deviation	Effect Size
Mother Gave Birth to Study Child as a Teen	0.1726202	0.1513973	0.0212229	1.1319239	0.2629606	0.368609	0.0575757
Race: Hispanic	0.3873536	0.4518326	-0.064479	-1.636733	0.1078432	0.493411	-0.13068
Peabody Picture Vocabulary Test (PPVT)	271.2665	271.64437	-0.377868	-0.137077	0.8915093	39.51132	-0.009564
Test de Vocabulario en Imágenes Peabody (TVIP)	264.02674	272.35775	-8.331009	-2.091674	0.0414662	41.735	-0.199617
Child's Primary Language at Baseline	0.7192439	0.6488639	0.07038	1.9853813	0.0524918	0.464676	0.1514604
Mother Separated/Divorced/Widowed	0.1827295	0.1558551	0.0268745	0.9924565	0.3256605	0.375193	0.0716284
Mother's Education-High School	0.295796	0.3127187	-0.016923	-0.622705	0.536253	0.460037	-0.036786
Mother's Education-Less than High School	0.3883955	0.4289865	-0.040591	-1.436105	0.1570783	0.491534	-0.08258
Mother's Education-More than High School	0.3158085	0.2582948	0.0575137	1.8228407	0.0741919	0.452595	0.1270756
Mother Married	0.4572724	0.4823749	-0.025103	-0.772111	0.4436121	0.499077	-0.050298
Mother Not Married	0.3599981	0.36177	-0.001772	-0.050415	0.959989	0.480253	-0.00369
WJ III Applied Problems	391.94686	388.60944	3.3374234	2.0381062	0.0467415	23.49321	0.1420591
WJ III Letter-Word Identification	309.18307	305.583	3.600063	1.6013755	0.1154705	25.65244	0.14034
WM Problemas Aplicados	396.80777	403.51372	-6.705958	-2.558092	0.0135386	22.32958	-0.300317
WM Identificación de letras y palabras	353.33954	358.82926	-5.489713	-2.199865	0.0323736	13.63738	-0.402549
Race: White	0.3490116	0.3025569	0.0464547	1.3114726	0.1955718	0.4688	0.0990926

Exhibit B.1D. Covariate Comparison of 3rd Grade Sample: 4-Year-Old Cohort (Principal Survey)

Variable	Treatment (T)	Control (C)	Difference (T-C)	t-test	p-value	Standard Deviation	Effect Size
Race: Black	0.2648296	0.2199738	0.0453306	1.4007576	0.1673419	0.428842	0.1057046
Caregiver Age as of 9/1/2002	29.10931	29.399293	-0.22868	-0.507436	0.6140363	6.928722	-0.033005
Family Cultural Enrichment Scale	3.4168853	3.5080751	-0.086458	-0.779274	0.4394197	1.353765	-0.063865
Parent Read to Child in Last Week	0.3850723	0.3190011	0.0632374	1.6526098	0.1045544	0.477837	0.1323409
Parental Safety Practices Scale	3.6544228	3.6563418	-0.00404	-0.147687	0.8831719	0.351767	-0.011485
Parent Spanked Child in Last Week	0.3993369	0.4176811	-0.015435	-0.488532	0.6272673	0.491519	-0.031404
Parent Used Time Out in Last Week	0.5986597	0.6002789	-0.003515	-0.082306	0.9347252	0.49001	-0.007174
Both Biological Parents Live with Child	0.4954291	0.5630214	-0.065598	-1.689548	0.0972197	0.49919	-0.13141
Child Age at Time of Spring Assessment	468.31518	468.69692	-0.355682	-0.215097	0.8305492	19.212	-0.018514
Number of Weeks Between 9/1/2002 and Date of Testing	241.17631	241.12985	0.0118309	0.0597716	0.9525711	2.981128	0.0039686
Child Gender	0.5017942	0.4749766	0.027413	0.615532	0.5409407	0.499872	0.05484
Primary Language Spoken at Home	0.6559032	0.6409932	0.0139022	0.3268735	0.7451014	0.477402	0.0291204
Biological Mother is a Recent Immigrant	0.2304091	0.2518622	-0.020717	-0.573723	0.5686771	0.427625	-0.048447
Aggressive Behavior	2.9640023	3.0676518	-0.106483	-0.863155	0.3920933	1.746946	-0.060954
Child Received Dental Care	0.8158366	0.657921	0.1565752	3.1234811	0.0029448	0.439357	0.3563732
Child's Overall Health Status is Excellent/Good	0.7999701	0.8306056	-0.027195	-1.121538	0.2673121	0.388347	-0.070029
Child Had Care for Injury Last Month	0.0479805	0.0757078	-0.02745	-1.672751	0.1005008	0.240298	-0.114232
Child Has Health Insurance Coverage	0.8622868	0.9017001	-0.039574	-1.345737	0.1843397	0.323146	-0.122466
Child Needs Ongoing Care	0.1336298	0.1727696	-0.042386	-1.579769	0.1203426	0.359751	-0.117821
Hyperactive Behavior	1.8530694	2.0055918	-0.150874	-1.537985	0.1302341	1.505812	-0.100194
Social Competencies	10.785549	10.845709	-0.053311	-0.421655	0.6750485	1.412261	-0.037749
Social Skills and Positive Approaches to Learning	12.310516	12.316185	0.0015111	0.0120184	0.9904579	1.734062	0.0008714
Total Child Behavior Problems	6.0893781	6.4044668	-0.316921	-1.142968	0.258389	3.712831	-0.085358
Number of Weeks Between 9/1/2002 and Parent Interview	.	.	.	.	.	.	.
Withdrawn Behavior	0.704359	0.800172	-0.102263	-1.420773	0.1614681	1.001244	-0.102136

Exhibit B.1D. Covariate Comparison of 3rd Grade Sample: 4-Year-Old Cohort (Principal Survey) (continued)

Variable	Treatment (T)	Control (C)	Difference (T-C)	t-test	p-value	Standard Deviation	Effect Size
Mother Gave Birth to Study Child as a Teen	0.1802379	0.1542723	0.026313	1.1175212	0.2690084	0.373466	0.0704562
Race: Hispanic	0.3990851	0.4545458	-0.054088	-1.239493	0.2208364	0.494521	-0.109374
Peabody Picture Vocabulary Test (PPVT)	270.71618	270.45784	0.1787405	0.0577255	0.9541929	40.21416	0.0044447
Test de Vocabulario en Imágenes Peabody (TVIP)	261.94842	273.78284	-11.83443	-3.543593	0.0008548	41.27054	-0.286752
Child's Primary Language at Baseline	0.6960464	0.6440031	0.0509453	1.2402499	0.2205589	0.46999	0.1083967
Mother Separated/Divorced/Widowed	0.1884192	0.1731415	0.0157043	0.5780867	0.5657495	0.38498	0.0407926
Mother's Education-High School	0.3221808	0.3278604	-0.007774	-0.223734	0.8238584	0.468358	-0.016599
Mother's Education-Less than High School	0.3897708	0.4470107	-0.055883	-1.805541	0.076894	0.493187	-0.113309
Mother's Education-More than High School	0.2880484	0.2251289	0.0636568	1.5957235	0.1167294	0.437148	0.1456184
Mother Married	0.4318363	0.4968552	-0.06321	-1.719569	0.091575	0.498674	-0.126756
Mother Not Married	0.3797445	0.3300033	0.0475058	1.1412509	0.2590961	0.478646	0.0992503
WJ III Applied Problems	394.0664	387.30312	6.4914972	3.6623878	0.0005942	22.67392	0.2862979
WJ III Letter-Word Identification	309.16941	304.02964	4.8577227	2.2819465	0.0266974	25.54613	0.1901549
WM Problemas Aplicados	397.0518	402.32806	-5.276259	-1.70629	0.0940374	22.93062	-0.230097
WM Identificación de letras y palabras	352.63941	358.50448	-5.865071	-2.503748	0.0155303	13.60077	-0.431231
Race: White	0.3360853	0.3254805	0.0087573	0.221275	0.8257619	0.470538	0.0186111

Exhibit B.1E. Covariate Comparison of 3rd Grade Sample: 4-Year-Old Cohort (Reading Teacher Survey-Teacher Child Report)

Variable	Treatment (T)	Control (C)	Difference (T-C)	t-test	p-value	Standard Deviation	Effect Size
Race: Black	0.2832574	0.2338604	0.049894	1.6092088	0.1137441	0.43814	0.1138767
Caregiver Age as of 9/1/2002	28.957288	28.933108	0.0832295	0.1635681	0.8707177	6.588734	0.0126321
Family Cultural Enrichment Scale	3.4802671	3.5264382	-0.041521	-0.313226	0.7553863	1.374194	-0.030215
Parent Read to Child in Last Week	0.4107914	0.3279234	0.0800262	1.7905611	0.0793001	0.482875	0.1657285
Parental Safety Practices Scale	3.6598268	3.6577878	-0.000081	-0.002906	1	0.356281	-0.000228
Parent Spanked Child in Last Week	0.4011594	0.4434577	-0.039267	-1.11199	0.271357	0.493854	-0.079512
Parent Used Time Out in Last Week	0.5992184	0.6099308	-0.012551	-0.342706	0.7332281	0.488967	-0.025668
Both Biological Parents Live with Child	0.4786668	0.5378014	-0.057237	-1.260561	0.2132037	0.499942	-0.114488
Child Age at Time of Spring Assessment	467.98516	469.98713	-1.96776	-1.088282	0.2815867	18.74594	-0.10497
Number of Weeks Between 9/1/2002 and Date of Testing	241.30095	241.16166	0.1048267	0.4579693	0.6489199	2.974344	0.0352436
Child Gender	0.4924906	0.4884978	0.0046817	0.1112561	0.91185	0.499911	0.0093651
Primary Language Spoken at Home	0.6727235	0.6492747	0.022446	0.5304467	0.5981045	0.473283	0.0474263
Biological Mother is a Recent Immigrant	0.2263461	0.25392	-0.02682	-0.597503	0.5528156	0.42698	-0.062813
Aggressive Behavior	2.8853524	3.0167611	-0.134296	-1.086994	0.2821503	1.755588	-0.076496
Child Received Dental Care	0.8006765	0.6814784	0.1180291	2.4839877	0.016318	0.437323	0.2698901
Child's Overall Health Status is Excellent/Good	0.8151476	0.8256127	-0.007104	-0.260924	0.7952018	0.383966	-0.018502
Child Had Care for Injury Last Month	0.0562251	0.0450702	0.0112463	0.8584017	0.3946872	0.219527	0.0512297
Child Has Health Insurance Coverage	0.8798212	0.8980683	-0.018472	-0.568405	0.5722545	0.314448	-0.058746
Child Needs Ongoing Care	0.1361711	0.1770813	-0.044128	-1.4725	0.1470308	0.363025	-0.121558
Hyperactive Behavior	1.87189	2.0210276	-0.147466	-1.276787	0.2074605	1.535858	-0.096015
Social Competencies	10.788284	10.90014	-0.104748	-0.824508	0.4134908	1.424495	-0.073533
Social Skills and Positive Approaches to Learning	12.320084	12.36754	-0.040054	-0.320873	0.749618	1.659108	-0.024142
Total Child Behavior Problems	6.0231618	6.3154624	-0.294424	-0.983913	0.3298031	3.791876	-0.077646
Number of Weeks Between 9/1/2002 and Parent Interview	.	.	.	.	.	.	.
Withdrawn Behavior	0.7042112	0.7836855	-0.085987	-1.271265	0.2094019	1.011791	-0.084985

Exhibit B.1E. Covariate Comparison of 3rd Grade Sample: 4-Year-Old Cohort (Reading Teacher Survey-Teacher Child Report) (continued)

Variable	Treatment (T)	Control (C)	Difference (T-C)	t-test	p-value	Standard Deviation	Effect Size
Mother Gave Birth to Study Child as a Teen	0.1781683	0.1517904	0.0267155	1.0542056	0.2967577	0.371422	0.0719275
Race: Hispanic	0.3791185	0.4338827	-0.053456	-1.268801	0.2102725	0.491066	-0.108857
Peabody Picture Vocabulary Test (PPVT)	270.21178	272.21398	-2.068028	-0.760776	0.4502942	40.17818	-0.051471
Test de Vocabulario en Imágenes Peabody (TVIP)	263.50655	275.58891	-12.08235	-3.266128	0.0019526	43.58376	-0.277221
Child's Primary Language at Baseline	0.7193454	0.6660237	0.0522851	1.2961111	0.2007721	0.461137	0.1133829
Mother Separated/Divorced/Widowed	0.2023197	0.1744546	0.0282596	0.9104633	0.3668614	0.391264	0.0722265
Mother's Education-High School	0.2987139	0.3250142	-0.02834	-0.845575	0.4017399	0.463138	-0.061191
Mother's Education-Less than High School	0.3832618	0.4246634	-0.040153	-1.311498	0.1955634	0.490601	-0.081845
Mother's Education-More than High School	0.3180243	0.2503224	0.0684931	1.8390314	0.0717358	0.451374	0.1517438
Mother Married	0.4288871	0.4814314	-0.050816	-1.362903	0.1789007	0.497933	-0.102053
Mother Not Married	0.3687932	0.344114	0.0225562	0.591315	0.5569213	0.479032	0.047087
WJ III Applied Problems	393.42946	387.58699	5.5852583	2.5162156	0.0150512	23.10474	0.2417364
WJ III Letter-Word Identification	309.77907	304.37102	5.1305954	2.57857	0.0128504	25.51113	0.2011121
WM Problemas Aplicados	395.7031	403.39298	-7.689875	-1.921407	0.0602775	23.31053	-0.329889
WM Identificación de letras y palabras	352.72533	359.00242	-6.27709	-2.409645	0.0196152	14.01538	-0.447872
Race: White	0.3376241	0.3322569	0.0035617	0.1079984	0.9144208	0.47199	0.0075462

Exhibit B.1F. Covariate Comparison of 3rd Grade Sample: 4-Year-Old Cohort (General Teacher Survey-Teacher Child Report)

Variable	Treatment (T)	Control (C)	Difference (T-C)	t-test	p-value	Standard Deviation	Effect Size
Race: Black	0.2695305	0.2304117	0.0396208	1.4167223	0.1626438	0.433158	0.0914698
Caregiver Age as of 9/1/2002	29.17386	29.102219	0.1313558	0.2535474	0.8008646	6.713182	0.0195669
Family Cultural Enrichment Scale	3.4458389	3.5205863	-0.070005	-0.566594	0.5734754	1.377248	-0.050829
Parent Read to Child in Last Week	0.3980907	0.3226247	0.0726528	1.8312168	0.0729126	0.480261	0.1512777
Parental Safety Practices Scale	3.6604265	3.6617508	-0.003423	-0.122371	0.9030864	0.354668	-0.009652
Parent Spanked Child in Last Week	0.4030981	0.4422885	-0.036173	-1.0259	0.3097801	0.493944	-0.073233
Parent Used Time Out in Last Week	0.5896014	0.6099008	-0.022101	-0.556763	0.5801244	0.489978	-0.045106
Both Biological Parents Live with Child	0.4981287	0.5429046	-0.042903	-1.048628	0.2992937	0.499592	-0.085876
Child Age at Time of Spring Assessment	468.10819	470.23648	-2.092607	-1.272226	0.2090629	18.76558	-0.111513
Number of Weeks Between 9/1/2002 and Date of Testing	241.37182	241.2307	0.1068616	0.4663583	0.642945	3.027206	0.0353004
Child Gender	0.493266	0.4869704	0.0069656	0.1584614	0.8747191	0.499903	0.0139338
Primary Language Spoken at Home	0.6483312	0.6496678	-0.002255	-0.064944	0.9484724	0.477286	-0.004725
Biological Mother is a Recent Immigrant	0.2397456	0.2551742	-0.014716	-0.30644	0.7605171	0.431472	-0.034106
Aggressive Behavior	2.9280033	2.9896284	-0.064769	-0.540476	0.5912213	1.741226	-0.037197
Child Received Dental Care	0.7945631	0.6834508	0.1099828	2.4234253	0.0189621	0.438743	0.2506772
Child's Overall Health Status is Excellent/Good	0.810073	0.8195968	-0.006188	-0.22185	0.8253167	0.388487	-0.015927
Child Had Care for Injury Last Month	0.0554353	0.0447981	0.0107268	0.8384908	0.405668	0.218343	0.0491282
Child Has Health Insurance Coverage	0.8780226	0.9016298	-0.023803	-0.77404	0.4424809	0.313316	-0.075971
Child Needs Ongoing Care	0.1431896	0.1876323	-0.047616	-1.639427	0.1072794	0.371265	-0.128254
Hyperactive Behavior	1.9158334	1.9859932	-0.06882	-0.691702	0.4922616	1.523267	-0.045179
Social Competencies	10.789762	10.914153	-0.117217	-0.994098	0.3248685	1.416644	-0.082743
Social Skills and Positive Approaches to Learning	12.288841	12.359691	-0.063424	-0.50882	0.6130727	1.681302	-0.037723
Total Child Behavior Problems	6.0613121	6.2326071	-0.173939	-0.681732	0.4984931	3.726137	-0.046681
Number of Weeks Between 9/1/2002 and Parent Interview	.	.	.	.	.	.	.
Withdrawn Behavior	0.6822071	0.7718484	-0.096132	-1.508578	0.1375771	0.992654	-0.096843

Exhibit B.1F. Covariate Comparison of 3rd Grade Sample: 4-Year-Old Cohort (General Teacher Survey-Teacher Child Report) (continued)

I

Variable	Treatment (T)	Control (C)	Difference (T-C)	t-test	p-value	Standard Deviation	Effect Size
Mother Gave Birth to Study Child as a Teen	0.1695189	0.1540688	0.0158143	0.612085	0.5432009	0.368363	0.0429314
Race: Hispanic	0.4020737	0.4410723	-0.037734	-1.016711	0.3140902	0.493767	-0.07642
Peabody Picture Vocabulary Test (PPVT)	269.2508	271.67618	-2.48842	-0.891384	0.3769094	39.98487	-0.062234
Test de Vocabulario en Imágenes Peabody (TVIP)	263.03287	275.22344	-12.19057	-3.588138	0.0007463	43.07901	-0.282982
Child's Primary Language at Baseline	0.6941568	0.6638344	0.0293568	0.8607766	0.3933899	0.466779	0.0628922
Mother Separated/Divorced/Widowed	0.2020577	0.1745621	0.0278815	0.8787808	0.3836418	0.391116	0.0712871
Mother's Education-High School	0.272395	0.3185532	-0.048153	-1.340885	0.1858997	0.456107	-0.105574
Mother's Education-Less than High School	0.4168297	0.4355976	-0.017571	-0.499949	0.6192613	0.494506	-0.035533
Mother's Education-More than High School	0.3107753	0.2458492	0.0657245	1.7857858	0.0800802	0.448397	0.1465765
Mother Married	0.435136	0.4825362	-0.045679	-1.261256	0.2129556	0.498275	-0.091675
Mother Not Married	0.3628064	0.3429016	0.0177977	0.4897865	0.6263853	0.477902	0.0372413
WJ III Applied Problems	393.28736	387.76662	5.2668182	2.421029	0.0190742	23.12048	0.2277989
WJ III Letter-Word Identification	309.22082	304.50123	4.4454675	2.0633123	0.0441908	25.71234	0.1728924
WM Problemas Aplicados	395.89346	403.54517	-7.651707	-2.296283	0.0258022	23.38402	-0.32722
WM Identificación de letras y palabras	352.80442	358.83453	-6.030115	-2.3669	0.0217716	13.83169	-0.435964
Race: White	0.3283958	0.328516	-0.001887	-0.058905	0.9532583	0.469651	-0.004018

Exhibit B.2A. Covariate Comparison of 3rd Grade Sample: 3-Year-Old Cohort (Child Assessment)

Variable	Treatment (T)	Control (C)	Difference (T-C)	t-test	p-value	Standard Deviation	Effect Size
Race: Black	0.3457307	0.3416624	0.0040683	0.199471	0.8426873	0.474943	0.008566
Caregiver Age as of 9/1/2002	29.622231	28.272716	1.3495157	2.680042	0.0098879	7.196336	0.187528
Family Cultural Enrichment Scale	3.4257697	3.2718812	0.1538885	1.62038	0.1113183	1.332139	0.11552
Parent Read to Child in Last Week	0.3613471	0.3153441	0.046003	1.656858	0.1036885	0.473164	0.097224
Parental Safety Practices Scale	3.6779168	3.6355427	0.0423741	1.425736	0.1600369	0.361081	0.117353
Parent Spanked Child in Last Week	0.4602351	0.5084503	-0.048215	-1.820779	0.0745096	0.499753	-0.096478
Parent Used Time Out in Last Week	0.6031306	0.6549166	-0.051786	-1.47375	0.1466949	0.483081	-0.107199
Both Biological Parents Live with Child	0.5047728	0.5018497	0.0029232	0.097961	0.9223471	0.499989	0.005847
Child Age at Time of Spring Assessment	474.49521	473.02696	1.4682433	1.573358	0.1218198	17.52155	0.083796
Number of Weeks Between 9/1/2002 and Date of Testing	293.04045	292.9696	0.0708453	0.437297	0.6637419	2.839803	0.024947
Child Gender	0.5112307	0.5275414	-0.016311	-0.437339	0.6637121	0.499625	-0.032646
Primary Language Spoken at Home	0.7433293	0.7380481	0.0052812	0.252142	0.8019446	0.438254	0.012051
Biological Mother is a Recent Immigrant	0.1654949	0.1315347	0.0339602	1.89242	0.0641178	0.355646	0.095489
Aggressive Behavior	3.1005915	2.9936321	0.1069594	0.905531	0.3694423	1.72224	0.062105
Child Received Dental Care	0.7565667	0.5889504	0.1676163	5.818175	3.9467E-07	0.469139	0.357285
Child's Overall Health Status is Excellent/Good	0.7983672	0.7813089	0.0170583	0.666068	0.5083697	0.40741	0.04187
Child Had Care for Injury Last Month	0.0608916	0.0698201	-0.008929	-0.649961	0.5186346	0.247136	-0.036128
Child Has Health Insurance Coverage	0.9079841	0.8730211	0.034963	1.609411	0.1136999	0.312215	0.111984
Child Needs Ongoing Care	0.178974	0.2492281	-0.070254	-3.364138	0.0014642	0.410144	-0.171291
Hyperactive Behavior	1.8529886	1.9210837	-0.068095	-0.640728	0.5245684	1.539654	-0.044228
Social Competencies	10.721949	10.6355	0.0864491	0.976234	0.3335566	1.552053	0.0557
Social Skills and Positive Approaches to Learning	12.191624	12.05921	0.1324142	1.031655	0.3071017	1.815366	0.072941
Total Child Behavior Problems	6.0998057	6.045396	0.0544097	0.209979	0.8345205	3.565546	0.01526
Number of Weeks Between 9/1/2002 and Parent Interview	.	.	.	.	.	.	.

Exhibit B.2A. Covariate Comparison of 3rd Grade Sample: 3-Year-Old Cohort (Child Assessment) (continued)

Variable	Treatment (T)	Control (C)	Difference (T-C)	t-test	p-value	Standard Deviation	Effect Size
Withdrawn Behavior	0.630421	0.617242	0.0131789	0.192993	0.8477308	0.907277	0.014526
Mother Gave Birth to Study Child as a Teen	0.1326439	0.1649821	-0.032338	-1.328135	0.1900468	0.355869	-0.090871
Race: Hispanic	0.3500891	0.3281125	0.0219766	1.072456	0.288564	0.473412	0.046422
Peabody Picture Vocabulary Test (PPVT)	229.87964	230.1417	-0.262061	-0.097508	0.9227052	36.96793	-0.007089
Test de Vocabulario en Imágenes Peabody (TVIP)	225.54728	241.8454	-16.29812	-2.890909	0.0056311	35.29303	-0.461794
Child's Primary Language at Baseline	0.7743646	0.7731754	0.0011893	0.05916	0.953056	0.418389	0.002843
Mother Separated/Divorced/Widowed	0.1162169	0.1344917	-0.018275	-0.851561	0.398439	0.331098	-0.055195
Mother's Education-High School	0.3591068	0.3142784	0.0448284	1.57136	0.1222831	0.472596	0.094856
Mother's Education-Less than High School	0.3139505	0.3506633	-0.036713	-1.324047	0.1913913	0.471026	-0.077942
Mother's Education-More than High School	0.3269427	0.3350583	-0.008116	-0.275369	0.7841452	0.47057	-0.017246
Mother Married	0.4630472	0.466614	-0.003567	-0.10481	0.9169378	0.498761	-0.007151
Mother Not Married	0.4157242	0.3988943	0.01683	0.523756	0.6027171	0.491337	0.034253
WJ III Applied Problems	367.35904	365.07983	2.2792178	1.039231	0.3035994	26.86318	0.084845
WJ III Letter-Word Identification	295.0802	292.82104	2.2591614	1.268907	0.2102351	21.94995	0.102923
WM Problemas Aplicados	384.49701	382.58178	1.9152265	0.900396	0.372142	23.77193	0.080567
WM Identificación de letras y palabras	352.5491	349.1044	3.4447023	2.171932	0.034535	12.11472	0.28434
Race: White	0.3041802	0.3302251	-0.026045	-1.147423	0.2565612	0.465376	-0.055965

Exhibit B.2B. Covariate Comparison of 3rd Grade Sample: 3-Year-Old Cohort (Math Teacher Survey-Teacher Child Report)

Variable	Treatment (T)	Control (C)	Difference (T-C)	t-test	p-value	Standard Deviation	Effect Size
Race: Black	0.3569218	0.3279263	0.0286565	1.0038	0.3202143	0.474545	0.060387
Caregiver Age as of 9/1/2002	29.707127	28.442931	1.2602511	2.583741	0.0126818	7.267951	0.173398
Family Cultural Enrichment Scale	3.3978741	3.2314827	0.1665236	1.587474	0.1185866	1.369301	0.121612
Parent Read to Child in Last Week	0.3501421	0.2998882	0.0496906	1.66676	0.1016928	0.468429	0.106079
Parental Safety Practices Scale	3.6510604	3.6245013	0.0268459	0.8666	0.3902201	0.370499	0.072459
Parent Spanked Child in Last Week	0.4668091	0.5171584	-0.050326	-1.717668	0.0919243	0.499934	-0.100666
Parent Used Time Out in Last Week	0.6077996	0.6570769	-0.048915	-1.584643	0.1192293	0.482176	-0.101447
Both Biological Parents Live with Child	0.4981259	0.4956894	0.0030303	0.081859	0.9350789	0.499991	0.006061
Child Age at Time of Spring Assessment	474.96843	472.21091	2.7846244	2.377576	0.021214	17.20792	0.161822
Number of Weeks Between 9/1/2002 and Date of Testing	292.83629	292.98099	-0.147443	-0.814515	0.4191369	2.741088	-0.05379
Child Gender	0.5168684	0.5136974	0.0031363	0.080518	0.9361408	0.499766	0.006276
Primary Language Spoken at Home	0.7455626	0.7254638	0.0200936	0.912161	0.3659759	0.441032	0.045561
Biological Mother is a Recent Immigrant	0.1775172	0.1321759	0.0452841	1.847059	0.0705437	0.361869	0.125139
Aggressive Behavior	3.1245561	2.9766264	0.1466675	1.220685	0.227819	1.716087	0.085466
Child Received Dental Care	0.7990823	0.5798655	0.2195192	8.273586	5.44E-11	0.462479	0.474658
Child's Overall Health Status is Excellent/Good	0.8077732	0.7753817	0.0325729	1.154689	0.2535996	0.406121	0.080205
Child Had Care for Injury Last Month	0.0469469	0.0612369	-0.014278	-1.143999	0.2579651	0.226127	-0.063142
Child Has Health Insurance Coverage	0.9157565	0.8692984	0.0464156	1.922891	0.0600864	0.309562	0.14994
Child Needs Ongoing Care	0.1716155	0.2622232	-0.090044	-3.390945	0.0013523	0.411988	-0.21856
Hyperactive Behavior	1.8667831	1.8399099	0.0289583	0.288209	0.7743547	1.524369	0.018997
Social Competencies	10.792915	10.63574	0.1553022	1.526932	0.1329565	1.551504	0.100098
Social Skills and Positive Approaches to Learning	12.341535	12.062356	0.2757619	2.452786	0.0176355	1.751185	0.157472
Total Child Behavior Problems	6.1631913	5.8861644	0.2782127	1.088936	0.2813012	3.59763	0.077332
Number of Weeks Between 9/1/2002 and Parent Interview	.	.	.	.	.	.	.

Exhibit B.2B. Covariate Comparison of 3rd Grade Sample: 3-Year-Old Cohort (Math Teacher Survey-Teacher Child Report) (continued)

Variable	Treatment (T)	Control (C)	Difference (T-C)	t-test	p-value	Standard Deviation	Effect Size
Withdrawn Behavior	0.6171475	0.5658716	0.0506201	0.708467	0.4818812	0.890362	0.056853
Mother Gave Birth to Study Child as a Teen	0.1232624	0.1695966	-0.046346	-1.646938	0.1057196	0.353418	-0.131136
Race: Hispanic	0.3460649	0.3353225	0.0113131	0.452468	0.6528509	0.473952	0.02387
Peabody Picture Vocabulary Test (PPVT)	230.99621	228.81169	2.1473106	0.756188	0.4530155	36.85032	0.058271
Test de Vocabulario en Imágenes Peabody (TVIP)	225.76707	242.72062	-16.95355	-3.245191	0.0020752	34.89234	-0.485882
Child's Primary Language at Baseline	0.7771798	0.7626905	0.014488	0.634436	0.5286321	0.42085	0.034426
Mother Separated/Divorced/Widowed	0.119842	0.1359499	-0.016523	-0.659679	0.5124281	0.333928	-0.04948
Mother's Education-High School	0.3285552	0.3066255	0.0222688	0.732241	0.4673734	0.465561	0.047832
Mother's Education-Less than High School	0.3439553	0.3574162	-0.013468	-0.403027	0.6886127	0.477174	-0.028225
Mother's Education-More than High School	0.3274894	0.3359583	-0.008801	-0.312273	0.7561063	0.470825	-0.018692
Mother Married	0.4569299	0.4551453	0.0022401	0.063149	0.9498948	0.498064	0.004498
Mother Not Married	0.4215836	0.4089048	0.0126402	0.327074	0.7449505	0.49277	0.025651
WJ III Applied Problems	368.36608	366.44543	1.9108591	0.867429	0.38977	27.16695	0.070338
WJ III Letter-Word Identification	295.227	292.75001	2.4739279	1.568067	0.1230498	22.11873	0.111848
WM Problemas Aplicados	384.12318	384.29142	-0.168237	-0.05938	0.9528814	23.80829	-0.007066
WM Identificación de letras y palabras	352.07669	349.1098	2.9668955	2.054827	0.0450357	11.71695	0.253214
Race: White	0.2970133	0.3367512	-0.03997	-1.691629	0.0968193	0.465221	-0.085915

Exhibit B.2C. Covariate Comparison of 3rd Grade Sample: 3-Year-Old Cohort (Parent Interview)

Variable	Treatment (T)	Control (C)	Difference (T-C)	t-test	p-value	Standard Deviation	Effect Size
Race: Black	0.3463493	0.3442124	0.002137	0.108374	0.9141245	0.47546	0.004495
Caregiver Age as of 9/1/2002	29.616115	28.20529	1.410825	2.943565	0.0048743	7.195384	0.196074
Family Cultural Enrichment Scale	3.4214788	3.2906487	0.1308301	1.361178	0.1794417	1.330939	0.098299
Parent Read to Child in Last Week	0.3573888	0.3178887	0.0395001	1.416194	0.1627977	0.472916	0.083525
Parental Safety Practices Scale	3.6730513	3.6370143	0.036037	1.295827	0.2008692	0.362947	0.09929
Parent Spanked Child in Last Week	0.4598623	0.5076667	-0.047804	-1.848249	0.0703684	0.499735	-0.095659
Parent Used Time Out in Last Week	0.6044576	0.6444884	-0.040031	-1.309545	0.1962187	0.484267	-0.082663
Both Biological Parents Live with Child	0.4970339	0.4952968	0.001737	0.059166	0.9530509	0.499985	0.003474
Child Age at Time of Spring Assessment	474.56847	473.13832	1.4301503	1.550198	0.127278	17.54884	0.081496
Number of Weeks Between 9/1/2002 and Date of Testing	.	.	.	.	.	.	.
Child Gender	0.511728	0.5217143	-0.009986	-0.281021	0.7798314	0.499721	-0.019984
Primary Language Spoken at Home	0.7456892	0.735648	0.0100411	0.481574	0.6321685	0.438263	0.022911
Biological Mother is a Recent Immigrant	0.1606292	0.1360337	0.0245955	1.36244	0.1790459	0.355449	0.069196
Aggressive Behavior	3.1208522	2.999436	0.1214162	1.103661	0.2749207	1.735979	0.069941
Child Received Dental Care	0.7522988	0.5905589	0.1617399	5.722736	5.5579E-07	0.469644	0.344388
Child's Overall Health Status is Excellent/Good	0.7972037	0.7795742	0.0176295	0.74084	0.4621879	0.40844	0.043163
Child Had Care for Injury Last Month	0.059749	0.0711233	-0.011374	-0.830527	0.410112	0.247277	-0.045998
Child Has Health Insurance Coverage	0.9110558	0.8707383	0.0403175	2.018272	0.0488373	0.311726	0.129336
Child Needs Ongoing Care	0.178289	0.2446492	-0.06636	-3.023725	0.0039019	0.408311	-0.162524
Hyperactive Behavior	1.8555247	1.9138183	-0.058294	-0.58919	0.558335	1.540221	-0.037848
Social Competencies	10.713385	10.667496	0.0458893	0.581917	0.5631859	1.560174	0.029413
Social Skills and Positive Approaches to Learning	12.214438	12.091271	0.1231672	0.992194	0.3257875	1.807859	0.068129
Total Child Behavior Problems	6.1429534	6.0403981	0.1025553	0.416315	0.6789268	3.602902	0.028465
Number of Weeks Between 9/1/2002 and Parent Interview	293.21576	293.12949	0.086263	0.519385	0.6057392	2.916021	0.029582

Exhibit B.2C. Covariate Comparison of 3rd Grade Sample: 3-Year-Old Cohort (Parent Interview) (continued)

Variable	Treatment (T)	Control (C)	Difference (T-C)	t-test	p-value	Standard Deviation	Effect Size
Withdrawn Behavior	0.6333202	0.6090704	0.0242498	0.35818	0.7216867	0.906685	0.026746
Mother Gave Birth to Study Child as a Teen	0.1310167	0.1653514	-0.034335	-1.376286	0.1747463	0.355254	-0.096648
Race: Hispanic	0.3451858	0.3280112	0.0171746	0.965933	0.3386359	0.472551	0.036344
Peabody Picture Vocabulary Test (PPVT)	229.91547	229.86112	0.0543424	0.022829	0.9818755	37.03351	0.001467
Test de Vocabulario en Imágenes Peabody (TVIP)	225.11162	241.68883	-16.57721	-3.029872	0.0038354	34.86992	-0.475401
Child's Primary Language at Baseline	0.7740846	0.770083	0.0040016	0.193582	0.8472722	0.419486	0.009539
Mother Separated/Divorced/Widowed	0.1204675	0.1346932	-0.014226	-0.652152	0.5172316	0.333609	-0.042642
Mother's Education-High School	0.3600568	0.3130578	0.0469989	1.715502	0.0923235	0.472546	0.099459
Mother's Education-Less than High School	0.3116733	0.3453724	-0.033699	-1.258223	0.214041	0.469666	-0.071751
Mother's Education-More than High School	0.3282699	0.3415698	-0.0133	-0.487641	0.6278937	0.471959	-0.02818
Mother Married	0.4504102	0.4658451	-0.015435	-0.442477	0.6600152	0.498243	-0.030979
Mother Not Married	0.424248	0.3994617	0.0247863	0.753202	0.4547919	0.492173	0.050361
WJ III Applied Problems	367.74189	364.78111	2.9607764	1.376714	0.1746147	26.88681	0.11012
WJ III Letter-Word Identification	294.98483	292.97271	2.0121156	1.201394	0.2351481	21.88804	0.091928
WM Problemas Aplicados	384.65981	383.23336	1.4264524	0.62918	0.5320394	23.35645	0.061073
WM Identificación de letras y palabras	352.52071	349.0909	3.4298124	2.23458	0.0298543	11.81299	0.290343
Race: White	0.3084649	0.3277764	-0.019312	-0.920694	0.3615452	0.46574	-0.041464

Exhibit B.2D. Covariate Comparison of 3rd Grade Sample: 3-Year-Old Cohort (Principal Survey)

Variable	Treatment (T)	Control (C)	Difference (T-C)	t-test	p-value	Standard Deviation	Effect Size
Race: Black	0.357764	0.3281	0.0293219	1.032714	0.3066102	0.4747	0.061769
Caregiver Age as of 9/1/2002	29.684527	28.681005	0.9998326	1.975891	0.0535891	7.307744	0.136818
Family Cultural Enrichment Scale	3.4185632	3.1696591	0.2489474	2.22586	0.0304703	1.355092	0.183713
Parent Read to Child in Last Week	0.3327155	0.3186148	0.0135752	0.435798	0.6648223	0.468629	0.028968
Parental Safety Practices Scale	3.6639018	3.6306685	0.0335115	1.051986	0.2977652	0.369106	0.090791
Parent Spanked Child in Last Week	0.4781075	0.4976604	-0.019565	-0.587219	0.5596478	0.499853	-0.039141
Parent Used Time Out in Last Week	0.6247629	0.656022	-0.030914	-0.94037	0.3514606	0.479896	-0.064419
Both Biological Parents Live with Child	0.5007259	0.5085473	-0.007215	-0.215139	0.8305166	0.499979	-0.014431
Child Age at Time of Spring Assessment	474.66064	472.73386	1.9551705	1.483105	0.1442003	17.72444	0.110309
Number of Weeks Between 9/1/2002 and Date of Testing	292.83436	292.88026	-0.048798	-0.274411	0.7848773	2.666764	-0.018299
Child Gender	0.5032971	0.5149543	-0.011679	-0.328579	0.7438193	0.499917	-0.023362
Primary Language Spoken at Home	0.7553456	0.7391074	0.0162379	0.660002	0.5122224	0.434592	0.037364
Biological Mother is a Recent Immigrant	0.1622279	0.135174	0.0270154	1.334322	0.1880259	0.355824	0.075924
Aggressive Behavior	3.1030902	2.9947434	0.1071217	0.853515	0.397365	1.726114	0.06206
Child Received Dental Care	0.7711605	0.5816541	0.1898421	6.44526	4.0914E-08	0.467766	0.405849
Child's Overall Health Status is Excellent/Good	0.7972917	0.777632	0.0198542	0.66528	0.5088692	0.409088	0.048533
Child Had Care for Injury Last Month	0.0501683	0.0682795	-0.018096	-1.282031	0.2056291	0.236005	-0.076676
Child Has Health Insurance Coverage	0.9136478	0.8627205	0.0508809	2.107394	0.0400201	0.315069	0.161491
Child Needs Ongoing Care	0.1604172	0.2457716	-0.084794	-3.053648	0.003588	0.402231	-0.210808
Hyperactive Behavior	1.8380772	1.8180964	0.0220801	0.211845	0.8330722	1.515877	0.014566
Social Competencies	10.749642	10.672417	0.0754307	0.661306	0.5113928	1.533149	0.0492
Social Skills and Positive Approaches to Learning	12.29565	12.044764	0.2474899	1.682911	0.0985054	1.839897	0.134513
Total Child Behavior Problems	6.0938557	5.8928534	0.202271	0.78764	0.4345529	3.587471	0.056383
Number of Weeks Between 9/1/2002 and Parent Interview	.	.	.	.	.	.	.

Exhibit B.2D. Covariate Comparison of 3rd Grade Sample: 3-Year-Old Cohort (Principal Survey) (continued)

Variable	Treatment (T)	Control (C)	Difference (T-C)	t-test	p-value	Standard Deviation	Effect Size
Withdrawn Behavior	0.5936378	0.5343048	0.0586676	0.782177	0.4377274	0.877245	0.066877
Mother Gave Birth to Study Child as a Teen	0.1209769	0.1523741	-0.031424	-1.223721	0.2266812	0.343467	-0.091489
Race: Hispanic	0.3307234	0.3365844	-0.005271	-0.19873	0.8432641	0.471515	-0.011179
Peabody Picture Vocabulary Test (PPVT)	231.46042	230.2867	1.137308	0.426035	0.6718749	37.28892	0.0305
Test de Vocabulario en Imágenes Peabody (TVIP)	226.4087	240.44213	-14.03343	-2.323837	0.0241564	35.71896	-0.392885
Child's Primary Language at Baseline	0.7852678	0.7723923	0.0128767	0.511113	0.6114775	0.415025	0.031026
Mother Separated/Divorced/Widowed	0.1228803	0.1405629	-0.018095	-0.719935	0.4748518	0.338166	-0.05351
Mother's Education-High School	0.3340352	0.3116744	0.0227005	0.794271	0.4307181	0.467577	0.048549
Mother's Education-Less than High School	0.3293836	0.3572534	-0.027863	-0.824937	0.4132493	0.474806	-0.058683
Mother's Education-More than High School	0.3365811	0.3310723	0.0051627	0.165818	0.8689558	0.471581	0.010948
Mother Married	0.4532438	0.4521055	0.0015962	0.040966	0.9674829	0.497755	0.003207
Mother Not Married	0.4225001	0.4073315	0.0151247	0.354151	0.724686	0.492711	0.030697
WJ III Applied Problems	367.93237	366.07412	1.8487184	0.814956	0.4188869	26.97827	0.068526
WJ III Letter-Word Identification	295.92046	292.90078	3.0159614	1.722012	0.0911279	21.86173	0.137956
WM Problemas Aplicados	382.7011	381.56078	1.1403288	0.437719	0.6634379	22.98937	0.049602
WM Identificación de letras y palabras	352.27489	349.78213	2.4927638	1.500446	0.1396649	12.87891	0.193554
Race: White	0.3115126	0.3353156	-0.024051	-0.994698	0.3245791	0.467769	-0.051417

Exhibit B.2E. Covariate Comparison of 3rd Grade Sample: 3-Year-Old Cohort (Reading Teacher Survey-Teacher Child Report)

Variable	Treatment (T)	Control (C)	Difference (T-C)	t-test	p-value	Standard Deviation	Effect Size
Race: Black	0.3499417	0.3374377	0.0121807	0.442628	0.6599062	0.474947	0.025647
Caregiver Age as of 9/1/2002	29.520753	28.303225	1.2136324	2.431909	0.0185698	7.090007	0.171175
Family Cultural Enrichment Scale	3.3743552	3.2694887	0.1050621	0.992567	0.3256069	1.352281	0.077693
Parent Read to Child in Last Week	0.3558386	0.3004787	0.05479	1.992504	0.051681	0.469582	0.116678
Parental Safety Practices Scale	3.6548616	3.6130992	0.0420334	1.224621	0.2263447	0.37482	0.112143
Parent Spanked Child in Last Week	0.4742028	0.5129368	-0.038724	-1.191219	0.239082	0.499958	-0.077455
Parent Used Time Out in Last Week	0.6159702	0.6751214	-0.058778	-2.002213	0.0505932	0.478383	-0.122868
Both Biological Parents Live with Child	0.5062153	0.4851977	0.0215936	0.605161	0.5477557	0.499982	0.043189
Child Age at Time of Spring Assessment	474.84766	472.41626	2.4589075	2.117804	0.0390869	17.24038	0.142625
Number of Weeks Between 9/1/2002 and Date of Testing	292.95226	292.90734	0.0419311	0.213996	0.8314037	2.769499	0.01514
Child Gender	0.5171927	0.5297332	-0.01256	-0.313486	0.75519	0.49945	-0.025149
Primary Language Spoken at Home	0.7413587	0.728989	0.0123734	0.539238	0.5920692	0.441228	0.028043
Biological Mother is a Recent Immigrant	0.1742044	0.1256773	0.0484663	2.018667	0.0487948	0.357106	0.13572
Aggressive Behavior	3.1356202	2.9978448	0.1365191	1.050061	0.2986404	1.721383	0.079308
Child Received Dental Care	0.776195	0.5744775	0.2020401	6.759048	1.3077E-08	0.468102	0.431616
Child's Overall Health Status is Excellent/Good	0.7970006	0.771856	0.0253341	0.924568	0.3595446	0.411184	0.061613
Child Had Care for Injury Last Month	0.052568	0.065159	-0.012581	-0.874446	0.3859747	0.235323	-0.053465
Child Has Health Insurance Coverage	0.912346	0.8750377	0.0372756	1.547822	0.1278489	0.308139	0.12097
Child Needs Ongoing Care	0.1775436	0.2534328	-0.075339	-2.79236	0.0073482	0.411058	-0.183281
Hyperactive Behavior	1.8744661	1.877879	-0.001292	-0.012186	0.9903249	1.531799	-0.000844
Social Competencies	10.740706	10.6382	0.100687	1.029686	0.3080163	1.572278	0.064039
Social Skills and Positive Approaches to Learning	12.289489	12.03492	0.2511709	2.16681	0.0349448	1.80268	0.139332
Total Child Behavior Problems	6.1695854	5.9443309	0.2264948	0.815204	0.4187463	3.613532	0.06268
Number of Weeks Between 9/1/2002 and Parent Interview	.	.	.	.	.	.	.

Exhibit B.2E. Covariate Comparison of 3rd grade Sample: 3-Year-Old Cohort (Reading Teacher Survey-Teacher Child Report) (continued)

Variable	Treatment (T)	Control (C)	Difference (T-C)	t-test	p-value	Standard Deviation	Effect Size
Withdrawn Behavior	0.6012512	0.5675368	0.0330755	0.456782	0.6497677	0.893309	0.037026
Mother Gave Birth to Study Child as a Teen	0.114066	0.1716276	-0.057561	-2.165899	0.0350181	0.349802	-0.164553
Race: Hispanic	0.348402	0.3339544	0.0150158	0.579748	0.5646369	0.474115	0.031671
Peabody Picture Vocabulary Test (PPVT)	230.39082	229.65378	0.7013087	0.265526	0.7916747	36.89586	0.019008
Test de Vocabulario en Imágenes Peabody (TVIP)	226.48801	243.13433	-16.64632	-3.120869	0.0029668	34.97943	-0.475889
Child's Primary Language at Baseline	0.7703036	0.7666113	0.0037032	0.14953	0.8817253	0.421814	0.008779
Mother Separated/Divorced/Widowed	0.1192528	0.1449491	-0.026102	-1.056664	0.2956445	0.338546	-0.077099
Mother's Education-High School	0.3584863	0.2958478	0.062935	1.854723	0.0694213	0.469224	0.134126
Mother's Education-Less than High School	0.3238185	0.3611169	-0.037281	-1.267922	0.2105837	0.474511	-0.078567
Mother's Education-More than High School	0.3176953	0.3430354	-0.025654	-0.837648	0.4061371	0.470327	-0.054546
Mother Married	0.4628276	0.4422649	0.0209998	0.585747	0.560629	0.497747	0.04219
Mother Not Married	0.4161718	0.412786	0.0033559	0.087682	0.9304727	0.492633	0.006812
WJ III Applied Problems	368.18801	366.25129	1.927013	0.828121	0.4114609	27.37221	0.0704
WJ III Letter-Word Identification	295.08077	292.60593	2.4717684	1.658394	0.1033768	22.16451	0.111519
WM Problemas Aplicados	384.8544	383.2506	1.6038025	0.625724	0.5342864	23.94312	0.066984
WM Identificación de letras y palabras	351.86532	348.4135	3.4518162	2.577172	0.0128964	11.34026	0.304386
Race: White	0.3016564	0.3286079	-0.027197	-1.064335	0.2921902	0.464548	-0.058544

Exhibit B.2F. Covariate Comparison of 3rd Grade Sample: 3-Year-Old Cohort (General Teacher Survey-Teacher Child Report)

Variable	Treatment (T)	Control (C)	Difference (T-C)	t-test	p-value	Standard Deviation	Effect Size
Race: Black	0.3367956	0.3312936	0.0051886	0.200882	0.8415898	0.471664	0.011001
Caregiver Age as of 9/1/2002	29.547791	28.365677	1.178249	2.53165	0.0144767	7.146095	0.16488
Family Cultural Enrichment Scale	3.3825469	3.2547557	0.1279698	1.249365	0.2172352	1.354697	0.094464
Parent Read to Child in Last Week	0.3475496	0.3064627	0.040529	1.440137	0.1559396	0.46919	0.086381
Parental Safety Practices Scale	3.6509219	3.6214737	0.0297357	0.872339	0.3871119	0.374462	0.079409
Parent Spanked Child in Last Week	0.4742459	0.5066595	-0.032405	-1.058457	0.2948346	0.499906	-0.064822
Parent Used Time Out in Last Week	0.6107127	0.660186	-0.049112	-1.590604	0.1178792	0.481369	-0.102025
Both Biological Parents Live with Child	0.5106608	0.5009476	0.0103006	0.295987	0.7684418	0.499966	0.020603
Child Age at Time of Spring Assessment	474.68192	471.94448	2.7647602	2.626931	0.01135	17.14781	0.161231
Number of Weeks Between 9/1/2002 and Date of Testing	292.90357	292.95623	-0.055528	-0.25569	0.7992184	2.78443	-0.019942
Child Gender	0.5185914	0.5146757	0.0038847	0.097055	0.9230633	0.499723	0.007774
Primary Language Spoken at Home	0.7299395	0.7179928	0.0119484	0.626221	0.5339631	0.447006	0.02673
Biological Mother is a Recent Immigrant	0.1841214	0.1295218	0.0545337	2.339199	0.0232802	0.363873	0.14987
Aggressive Behavior	3.1435882	2.9701423	0.1721534	1.330025	0.1894278	1.714961	0.100383
Child Received Dental Care	0.7797789	0.5901763	0.189937	7.001514	5.4127E-09	0.464171	0.409196
Child's Overall Health Status is Excellent/Good	0.8013364	0.7765687	0.0249597	0.889849	0.3777254	0.40797	0.06118
Child Had Care for Injury Last Month	0.0552444	0.0668194	-0.011566	-0.810303	0.421531	0.23929	-0.048334
Child Has Health Insurance Coverage	0.9138433	0.8716696	0.0421347	1.725791	0.0904397	0.309173	0.136282
Child Needs Ongoing Care	0.1759595	0.2532318	-0.07672	-2.731341	0.0086417	0.410291	-0.18699
Hyperactive Behavior	1.8818236	1.8217791	0.0621036	0.577346	0.5662457	1.536954	0.040407
Social Competencies	10.762214	10.667886	0.0925163	0.959086	0.3420405	1.553332	0.05956
Social Skills and Positive Approaches to Learning	12.300871	12.06757	0.2299217	1.880962	0.0656921	1.773982	0.129608
Total Child Behavior Problems	6.1871896	5.8386717	0.3496376	1.223419	0.2267943	3.615568	0.096703
Number of Weeks Between 9/1/2002 and Parent Interview	.	.	.	.	.	.	.

Exhibit B.2F. Covariate Comparison of 3rd Grade Sample: 3-Year-Old Cohort (General Teacher Survey-Teacher Child Report) (continued)

Variable	Treatment (T)	Control (C)	Difference (T-C)	t-test	p-value	Standard Deviation	Effect Size
Withdrawn Behavior	0.59873	0.5516671	0.0464111	0.621546	0.5370088	0.891276	0.052073
Mother Gave Birth to Study Child as a Teen	0.1137959	0.1781161	-0.064313	-2.387966	0.0206836	0.352761	-0.182314
Race: Hispanic	0.3705204	0.3424899	0.0285849	1.104975	0.2743564	0.479006	0.059676
Peabody Picture Vocabulary Test (PPVT)	230.06245	229.01059	1.0152829	0.401059	0.6900523	36.79586	0.027592
Test de Vocabulario en Imágenes Peabody (TVIP)	226.27818	242.75935	-16.48117	-3.083073	0.0033024	34.87928	-0.47252
Child's Primary Language at Baseline	0.7585649	0.7526818	0.0058896	0.262054	0.7943352	0.429701	0.013706
Mother Separated/Divorced/Widowed	0.1170658	0.1383771	-0.021726	-0.886208	0.379665	0.333668	-0.065112
Mother's Education-High School	0.3341971	0.3039324	0.0305956	1.068396	0.2903732	0.466169	0.065632
Mother's Education-Less than High School	0.3403065	0.3592575	-0.018949	-0.587678	0.5593418	0.476873	-0.039736
Mother's Education-More than High School	0.3254964	0.3368101	-0.011646	-0.395557	0.6940818	0.470609	-0.024747
Mother Married	0.4664851	0.4530607	0.0138688	0.388074	0.6995766	0.498384	0.027827
Mother Not Married	0.4148169	0.4085622	0.0062266	0.154967	0.8774588	0.492145	0.012652
WJ III Applied Problems	368.08976	366.397	1.6830897	0.664146	0.5095886	27.36565	0.061504
WJ III Letter-Word Identification	294.88605	292.84948	2.0338288	1.266745	0.2110011	22.02244	0.092353
WM Problemas Aplicados	384.61333	383.92883	0.6844953	0.281572	0.779411	23.85828	0.02869
WM Identificación de letras y palabras	352.18333	349.12569	3.057645	2.213593	0.0313558	11.56843	0.264309
Race: White	0.292684	0.3262164	-0.033774	-1.486325	0.1433495	0.462202	-0.073071

Appendix C

Intent to Treat (ITT) Tables, 2002-2008

Exhibit C.1A. Estimated Impacts on Cognitive Outcomes by Year: 4-Year-Old Cohort

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Head Start Baseline (Fall 2002)							
Language and Literacy Measures++							
PPVT (Adapted)	270.54	271.73					
WJ III Letter-Word Identification	309.63	306.73					
WJ III Spelling	359.73	356.41					
WJ III Oral Comprehension	442.92	445.43					
CTOPPP Elision	272.24	271.76					
Color Identification	0.58	0.51					
WJ III Pre-Academic Skills	357.81	357.34					
Spanish Language and Literacy Measures+++							
TVIP (Adapted)	263.08	270.47					
WM Letter-Word Identification	351.90	356.14					
Pre-writing Measure++							
McCarthy Draw-a-Design	3.88	3.77					
Parent-Reported Literacy Measure							
Emergent Literacy Scale	2.94	2.58					
Math Skills Measures++							
WJ III Applied Problems	391.58	389.55					
Counting Bears	0.41	0.39					
Head Start Year (Spring 2003)							
Language and Literacy Measures							
PPVT (Adapted)	294.35	290.25	4.10	0.060	**3.55****	0.028	0.09
WJ III Letter-Word Identification	325.46	319.22	6.24	0.034	**5.98****	0.017	0.22
WJ III Spelling	371.56	367.67	3.89	0.046	**3.77****	0.029	0.15
WJ III Oral Comprehension	443.40	443.65	-0.24	0.818	-0.94	0.395	-0.05
CTOPPP Elision	273.85	271.41	2.45	0.463	2.45	0.444	0.05
Color Identification	0.73	0.66	0.07	0.019	**0.08****	0.010	0.16
Letter Naming	11.53	9.21	2.33	0.008	**2.36*****	0.002	0.25
WJ III Pre-Academic Skills	365.00	360.56	4.45	0.041	**4.23****	0.022	0.19

Exhibit C.1A. Estimated Impacts on Cognitive Outcomes by Year: 4-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Spanish Language and Literacy Measures+++							
TVIP (Adapted)	298.54	290.77	7.77	0.380	9.04	0.106	0.21
WM Letter-Word Identification	360.70	359.23	1.47	0.328	1.91	0.180	0.14
Pre-writing Measure							
McCarthy Draw-a-Design	4.58	4.40	0.19	0.135	0.20	0.110	0.10
Parent-Reported Literacy Measure							
Emergent Literacy Scale	3.76	3.35	0.42	0.000	**0.43*** **	0.000	0.31
Math Skills Measures							
WJ III Applied Problems	397.47	394.42	3.05	0.178	3.17	0.139	0.12
Counting Bears	0.59	0.55	0.04	0.185	0.04	0.181	0.08
Kindergarten Year (Spring 2004)							
Language and Literacy Measures							
PPVT (Adapted)	334.21	331.85	2.37	0.398	1.78	0.328	0.04
WJ III Letter-Word Identification	378.08	378.15	-0.08	0.970	-0.19	0.918	-0.01
WJ III Spelling	413.91	414.12	-0.21	0.899	-0.52	0.764	-0.02
WJ III Oral Comprehension	456.52	457.29	-0.77	0.551	-0.91	0.327	-0.05
CTOPPP Elision	321.89	323.91	-2.02	0.586	-2.85	0.374	-0.06
Letter Naming	22.99	22.65	0.34	0.351	0.40	0.274	0.06
WJ III Pre-Academic Skills	406.23	406.48	-0.26	0.868	-0.47	0.745	-0.02
WJ III Word Attack	431.60	432.68	-1.09	0.628	-1.13	0.639	-0.03
WJ III Basic Reading Skills	404.79	405.39	-0.60	0.765	-0.71	0.728	-0.02
Spanish Language and Literacy Measures+++							
TVIP (Adapted)	326.48	327.18	-0.70	0.927	-1.03	0.868	-0.02
WM Letter-Word Identification	390.55	396.10	-5.55	0.062	-4.28	0.130	-0.16

Exhibit C.1A. Estimated Impacts on Cognitive Outcomes by Year: 4-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Math Skills Measures							
WJ III Applied Problems	426.59	426.32	0.27	0.872	0.12	0.936	0.01
WJ III Quantitative Concepts	441.83	441.88	-0.05	0.968	-0.13	0.920	-0.01
WJ III Math Reasoning	434.15	434.12	0.03	0.981	-0.07	0.951	0.00
School Performance Assessment Measures							
School Accomplishments	28.13	28.16	-0.03	0.969	0.00	0.997	0.00
Promotion	0.94	0.92	0.01	0.569	0.00	0.888	0.01
Language and Literacy Ability	0.75	0.71	0.04	0.381	0.04	0.424	0.08
Math Ability	0.80	0.75	0.05	0.164	0.05	0.191	0.11
Social Studies and Science Ability	0.83	0.80	0.03	0.433	0.03	0.501	0.07
1st Grade Year (Spring 2005)							
Language and Literacy Measures							
PPVT (Adapted)	363.07	358.74	4.34	0.075	2.95*	0.072	0.09
WJ III Letter-Word Identification	433.01	432.26	0.75	0.730	0.83	0.705	0.02
WJ III Spelling	451.88	450.13	1.76	0.312	1.55	0.347	0.06
WJ III Oral Comprehension	473.42	472.36	1.06	0.438	0.34	0.717	0.02
WJ III Pre-Academic Skills	446.66	445.44	1.22	0.412	0.95	0.510	0.04
WJ III Word Attack	469.10	467.41	1.69	0.344	1.71	0.324	0.05
WJ III Basic Reading Skills	451.04	449.81	1.22	0.521	1.08	0.550	0.03
WJ III Academic Applications	461.77	461.22	0.55	0.606	0.38	0.730	0.02
WJ III Academic Skills	449.02	447.71	1.30	0.380	1.11	0.446	0.05
WJ III Passage Comprehension	450.28	449.86	0.42	0.814	0.17	0.922	0.01
WJ III Writing Sample	479.87	479.75	0.12	0.863	0.15	0.824	0.01

Exhibit C.1A. Estimated Impacts on Cognitive Outcomes by Year: 4-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Spanish Language and Literacy Measures+++							
TVIP (Adapted)	376.86	372.20	4.65	0.361	5.25	0.240	0.13
WM Letter-Word Identification	407.84	415.07	-7.23	0.164	-4.30	0.397	-0.09
Math Skills Measures							
WJ III Applied Problems	455.16	454.13	1.03	0.405	0.81	0.523	0.04
WJ III Quantitative Concepts	461.79	461.28	0.51	0.714	0.31	0.819	0.02
WJ III Math Reasoning	458.36	457.67	0.68	0.580	0.47	0.705	0.03
WJ III Calculation	461.76	460.46	1.30	0.245	1.41	0.255	0.07
School Performance Assessment Measures							
School Accomplishments	43.25	43.79	-0.54	0.481	-0.59	0.500	-0.06
Promotion	0.93	0.92	0.01	0.323	0.01	0.376	0.05
Language and Literacy Ability	0.69	0.70	-0.01	0.705	-0.02	0.433	-0.05
Math Ability	0.77	0.81	-0.04	0.235	-0.05	0.148	-0.12
Social Studies and Science Ability	0.83	0.85	-0.02	0.449	-0.02	0.362	-0.06
3rd Grade Year (Spring 2007)							
Language and Literacy Measures							
ECLS-K Reading	98.61	96.63	1.98	0.139	2.23*	0.075	0.11
PPVT (Adapted)	408.14	405.74	2.40	0.298	2.17	0.246	0.08
WJ III Letter-Word Identification	482.10	480.60	1.51	0.450	2.11	0.275	0.07
Spanish Language and Literacy Measures+++							
WM Letter-Word Identification	464.78	462.31	2.47	0.787	3.53	0.678	0.07
Math Skills Measures							
WJ III Applied Problems	486.96	487.70	-0.74	0.601	-0.43	0.729	-0.02
WJ III Calculation	491.28	491.52	-0.24	0.826	0.00	0.997	0.00

Exhibit C.1A. Estimated Impacts on Cognitive Outcomes by Year: 4-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
School Performance Assessment Measures							
Promotion	0.96	0.95	0.00	0.885	0.01	0.768	0.02
Language and Literacy Ability	0.66	0.63	0.03	0.547	0.04	0.389	0.09
Math Ability	0.69	0.72	-0.03	0.454	-0.03	0.462	-0.07
Reading/Language Arts Skills++++	-0.06	-0.05	-0.02	0.855	-0.01	0.945	-0.01
Math Skills++++	-0.02	-0.01	-0.01	0.902	-0.03	0.632	-0.03

Key:

******* $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

++ Indicates baseline scores for English-speaking children only except for the PPVT and Woodcock-Johnson III Letter-Word test.

+++ The scores for the study children from Puerto Rico are not included in this analysis.

++++ These scores are reported in the Item Response Theory (IRT) metric. Since there is not a normalizing sample to link the scores to, the scores are reported in a standardized format with a mean of 0 and a standard deviation of 1.

Bold regression-adjusted impact value indicates the outcome passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.

Exhibit C.1B. Estimated Impacts on Cognitive Outcomes by Year: 3-Year-Old Cohort

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Head Start Baseline (Fall 2002)							
Language and Literacy Measures++							
PPVT (Adapted)	230.01	230.49					
WJ III Letter-Word Identification	294.70	293.38					
WJ III Spelling	334.81	333.54					
WJ III Oral Comprehension	433.32	433.72					
CTOPPP Elision	235.14	230.24					
Color Identification	0.29	0.27					
WJ III Pre-Academic Skills	336.81	335.88					
Spanish Language and Literacy Measures+++							
TVIP (Adapted)	223.02	236.01					
WM Letter-Word Identification	351.17	347.06					
Pre-writing Measure++							
McCarthy Draw-a-Design	2.72	2.70					
Parent-Reported Literacy Measure							
Emergent Literacy Scale+	2.05	1.93					
Math Skills Measures++							
WJ III Applied Problems	367.31	365.05					
Counting Bears	0.15	0.17					
Head Start Year (Spring 2003)							
Language and Literacy Measures							
PPVT (Adapted)	257.50	251.43	6.07	0.003	**6.53*** **	0.000	0.18
WJ III Letter-Word Identification	307.00	300.51	6.49	0.001	**6.14*** **	0.000	0.26
WJ III Spelling	346.57	343.64	2.93	0.061	2.28	0.130	0.10
WJ III Oral Comprehension	435.52	435.44	0.09	0.924	0.28	0.698	0.02
CTOPPP Elision	241.44	235.03	6.41	0.078	**5.01***	0.061	0.10
Color Identification	0.49	0.46	0.03	0.385	0.04	0.179	0.07
Letter Naming	5.49	3.92	1.57	0.010	**1.56*** **	0.005	0.24
WJ III Pre-Academic Skills	343.67	339.41	4.26	0.013	**4.25*** **	0.004	0.22

Exhibit C.1B. Estimated Impacts on Cognitive Outcomes by Year: 3-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Spanish Language and Literacy Measures+++							
TVIP (Adapted)	256.83	247.05	9.79	0.069	5.21	0.365	0.13
WM Letter-Word Identification	353.78	351.56	2.23	0.264	1.59	0.380	0.13
Pre-writing Measure							
McCarthy Draw-a-Design	3.23	3.05	0.18	0.005	**0.16*** **	0.007	0.14
Parent-Reported Literacy Measure							
Emergent Literacy Scale+	2.86	2.35	0.51	0.000	**0.48*** **	0.000	0.35
Math Skills Measures							
WJ III Applied Problems	377.27	373.57	3.69	0.144	4.31**	0.012	0.15
Counting Bears	0.30	0.28	0.02	0.362	0.03	0.241	0.06
Age 4 Year (Spring 2004)							
Language and Literacy Measures							
PPVT (Adapted)	300.99	298.28	2.72	0.177	2.03	0.251	0.05
WJ III Letter-Word Identification	333.08	330.13	2.95	0.112	2.56	0.112	0.09
WJ III Spelling	376.74	376.26	0.47	0.794	0.28	0.875	0.01
WJ III Oral Comprehension	446.19	445.80	0.38	0.662	0.25	0.743	0.02
CTOPPP Elision	281.07	271.90	9.17	0.004	**8.26*** **	0.002	0.15
Color Identification	0.81	0.80	0.02	0.432	0.01	0.466	0.03
Letter Naming	13.73	12.84	0.89	0.144	0.85	0.155	0.09
WJ III Pre-Academic Skills	370.24	368.72	1.53	0.300	1.24	0.378	0.06
Spanish Language and Literacy Measures+++							
TVIP (Adapted)	292.51	292.45	0.07	0.990	-1.33	0.803	-0.03
WM Letter-Word Identification	365.63	361.99	3.63	0.231	3.05	0.334	0.16
Pre-writing Measure							
McCarthy Draw-a-Design	4.83	4.94	-0.11	0.425	-0.09	0.482	-0.04
Parent-Reported Literacy Measure							
Emergent Literacy Scale	4.01	3.82	0.19	0.005	**0.20*** **	0.002	0.16

Exhibit C.1B. Estimated Impacts on Cognitive Outcomes by Year: 3-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Math Skills Measures							
WJ III Applied Problems	401.06	399.86	1.20	0.381	0.75	0.551	0.03
Counting Bears	0.56	0.55	0.01	0.879	0.01	0.777	0.02
Kindergarten Year (Spring 2005)							
Language and Literacy Measures							
PPVT (Adapted)	340.31	339.91	0.40	0.783	0.26	0.851	0.01
WJ III Letter-Word Identification	384.02	383.42	0.61	0.794	0.24	0.899	0.01
WJ III Spelling	420.20	419.43	0.77	0.670	0.45	0.774	0.02
WJ III Oral Comprehension	457.91	457.36	0.55	0.648	0.50	0.633	0.03
CTOPPP Elision	331.33	335.07	-3.74	0.220	-3.52	0.241	-0.08
Letter Naming	23.46	23.67	-0.21	0.580	-0.32	0.340	-0.06
WJ III Pre-Academic Skills	411.62	411.39	0.22	0.898	-0.02	0.988	0.00
WJ III Word Attack	436.00	437.37	-1.38	0.566	-1.37	0.563	-0.04
WJ III Basic Reading Skills	410.05	410.38	-0.33	0.883	-0.54	0.801	-0.02
Spanish Language and Literacy Measures+++							
TVIP (Adapted)	353.10	358.74	-5.64	0.210	-7.51	0.117	-0.19
WM Letter-Word Identification	393.41	382.28	11.13	0.009	8.73*	0.053	0.26
Math Skills Measures							
WJ III Applied Problems	430.39	431.29	-0.90	0.595	-0.94	0.519	-0.04
WJ III Quantitative Concepts	442.98	443.71	-0.73	0.497	-0.88	0.310	-0.05
WJ III Math Reasoning	436.69	437.50	-0.81	0.548	-0.91	0.408	-0.05

Exhibit C.1B. Estimated Impacts on Cognitive Outcomes by Year: 3-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
School Performance Assessment Measures							
School Accomplishments	27.57	28.32	-0.75	0.172	-0.65	0.203	-0.09
Promotion	0.89	0.90	-0.01	0.772	-0.01	0.709	-0.03
Language and Literacy Ability	0.75	0.79	-0.04	0.103	-0.04	0.127	-0.09
Math Ability	0.79	0.86	-0.07	0.004	**-0.07*****	0.003	-0.19
Social Studies and Science Ability	0.84	0.87	-0.03	0.171	-0.03	0.121	-0.10
1st Grade Year (Spring 2006)							
Language and Literacy Measures							
PPVT (Adapted)	360.41	357.91	2.50	0.280	2.32	0.151	0.08
WJ III Letter-Word Identification	433.69	432.92	0.78	0.729	0.37	0.848	0.01
WJ III Spelling	453.89	454.94	-1.04	0.548	-1.20	0.438	-0.05
WJ III Oral Comprehension	472.60	471.25	1.36	0.118	1.35*	0.051	0.08
WJ III Pre-Academic Skills	447.53	447.17	0.36	0.830	0.24	0.869	0.01
WJ III Word Attack	468.84	469.12	-0.28	0.890	-0.60	0.759	-0.02
WJ III Basic Reading Skills	451.29	450.97	0.32	0.877	-0.08	0.966	0.00
WJ III Academic Applications	463.05	462.29	0.76	0.524	0.73	0.489	0.04
WJ III Academic Skills	449.89	450.19	-0.30	0.844	-0.60	0.633	-0.03
WJ III Passage Comprehension	451.10	450.18	0.92	0.549	0.76	0.580	0.03
WJ III Writing Sample+	483.03	483.04	-0.01	0.990	-0.09	0.928	-0.01
Spanish Language and Literacy Measures+++							
TVIP (Adapted)	376.07	374.41	1.66	0.735	0.04	0.993	0.00
WM Letter-Word Identification	416.96	418.67	-1.71	0.739	-0.54	0.910	-0.01

Exhibit C.1B. Estimated Impacts on Cognitive Outcomes by Year: 3-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Math Skills Measures							
WJ III Applied Problems	455.07	453.67	1.40	0.292	1.59	0.161	0.08
WJ III Quantitative Concepts	462.01	461.36	0.64	0.615	0.79	0.450	0.05
WJ III Math Reasoning	458.49	457.45	1.04	0.405	1.20	0.231	0.07
WJ III Calculation	461.81	461.65	0.15	0.872	-0.02	0.977	0.00
School Performance Assessment Measures							
School Accomplishments	42.46	42.74	-0.28	0.728	-0.29	0.705	-0.03
Promotion	0.92	0.93	-0.01	0.480	-0.02	0.248	-0.07
Language and Literacy Ability	0.72	0.71	0.00	0.920	0.00	0.888	0.01
Math Ability	0.78	0.80	-0.02	0.450	-0.02	0.448	-0.05
Social Studies and Science Ability	0.83	0.86	-0.03	0.198	-0.03	0.286	-0.07
3rd Grade Year (Spring 2008)							
Language and Literacy Measures							
ECLS-K Reading	98.10	97.91	0.20	0.868	-0.18	0.876	-0.01
PPVT (Adapted)	407.85	405.67	2.18	0.122	1.83	0.146	0.06
WJ III Letter-Word Identification	483.60	482.81	0.79	0.661	0.44	0.818	0.01
Spanish Language and Literacy Measures+++							
WM Letter-Word Identification	468.63	470.77	-2.14	0.734	-1.63	0.804	-0.03
Math Skills Measures							
WJ III Applied Problems	486.44	486.48	-0.05	0.975	0.03	0.985	0.00
WJ III Calculation	491.79	491.66	0.13	0.896	-0.05	0.960	0.00

Exhibit C.1B. Estimated Impacts on Cognitive Outcomes by Year: 3-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
School Performance Assessment Measures							
Promotion	0.94	0.95	-0.01	0.332	-0.02*	0.092	-0.11
Language and Literacy Ability	0.70	0.73	-0.03	0.434	-0.04	0.372	-0.09
Math Ability	0.74	0.74	0.00	0.920	0.01	0.680	0.03
Reading/Language Arts Skills++++	0.04	0.06	-0.02	0.822	-0.05	0.515	-0.05
Math Skills++++	0.06	0.02	0.04	0.453	0.01	0.861	0.01

Key:

*** $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

++ Indicates baseline scores for English-speaking children only except for the PPVT and Woodcock-Johnson III Letter-Word test.

+++ The scores for the study children from Puerto Rico are not included in this analysis.

++++ These scores are reported in the Item Response Theory (IRT) metric. Since there is not a normalizing sample to link the scores to, the scores are reported in a standardized format with a mean of 0 and a standard deviation of 1.

Bold regression-adjusted impact value indicates the outcome passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.

Exhibit C.2A. Estimated Impacts on Social-Emotional Outcomes by Year: 4-Year-Old Cohort

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Head Start Baseline (Fall 2002)							
Parent-Reported Measures							
Aggressive Behavior	2.93	3.07					
Hyperactive Behavior	1.85	2.08					
Withdrawn Behavior	0.70	0.75					
Total Problem Behavior	6.06	6.43					
Social Competencies	10.78	10.84					
Social Skills and Positive Approaches to Learning	12.33	12.22					
Head Start Year (Spring 2003)							
Parent-Reported Measures							
Aggressive Behavior+	2.73	2.86	-0.13	0.263	-0.16	0.164	-0.10
Hyperactive Behavior+	1.71	1.77	-0.06	0.502	-0.09	0.324	-0.06
Withdrawn Behavior+	0.65	0.70	-0.04	0.505	-0.04	0.575	-0.04
Total Problem Behavior	5.60	5.80	-0.20	0.406	-0.27	0.289	-0.08
Social Competencies+	11.01	11.06	-0.04	0.665	-0.04	0.566	-0.03
Social Skills and Positive Approaches to Learning	12.46	12.48	-0.02	0.891	-0.06	0.682	-0.04
Closeness	33.58	33.31	0.27	0.097	0.25	0.146	0.09
Conflict	17.46	17.71	-0.25	0.683	-0.23	0.698	-0.03
Positive Relationships+	64.05	63.47	0.58	0.411	0.56	0.419	0.07
Kindergarten Year (Spring 2004)							
Parent-Reported Measures							
Aggressive Behavior+	2.41	2.47	-0.06	0.608	-0.08	0.477	-0.05
Hyperactive Behavior	1.53	1.39	0.15	0.174	0.11	0.273	0.08
Withdrawn Behavior+	0.74	0.71	0.02	0.721	0.00	0.986	0.00

Exhibit C.2A. Estimated Impacts on Social-Emotional Outcomes by Year: 4-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Parent-Reported Measures (cont'd)							
Total Problem Behavior	5.18	4.99	0.19	0.464	0.09	0.710	0.03
Social Competencies+	11.10	11.17	-0.07	0.382	-0.03	0.770	-0.02
Social Skills and Positive Approaches to Learning+	12.66	12.63	0.03	0.780	0.07	0.483	0.05
Closeness	33.19	33.34	-0.15	0.526	-0.06	0.793	-0.02
Conflict	17.68	17.59	0.09	0.854	-0.13	0.788	-0.02
Positive Relationships+	63.38	63.65	-0.27	0.643	0.03	0.956	0.00
Teacher-Reported Measures							
ASPI-Aggressive	48.74	48.72	0.02	0.973	-0.09	0.893	-0.01
ASPI-Inattentive/ Hyperactive	50.49	50.97	-0.48	0.468	-0.69	0.286	-0.08
ASPI-Withdrawn/ Low Energy	49.22	49.08	0.15	0.824	0.10	0.888	0.01
ASPI-Oppositional	48.03	47.67	0.37	0.556	0.13	0.819	0.02
ASPI-Problems with Peer Interaction	51.16	51.70	-0.54	0.629	-0.89	0.410	-0.08
ASPI-Shy/Socially Reticent	47.81	47.13	0.68	0.390	0.64	0.418	0.08
ASPI-Problems with Structured Learning	50.86	51.26	-0.40	0.623	-0.67	0.410	-0.07
ASPI-Problems with Teacher Interaction	50.07	49.79	0.28	0.729	0.20	0.811	0.02
Closeness	30.33	30.11	0.21	0.631	0.26	0.557	0.06
Conflict	13.33	13.52	-0.19	0.759	-0.35	0.558	-0.06
Positive Relationships	64.81	64.38	0.42	0.613	0.63	0.445	0.07
1st Grade Year (Spring 2005)							
Parent-Reported Measures							
Aggressive Behavior	2.20	2.29	-0.09	0.476	-0.09	0.483	-0.05
Hyperactive Behavior	1.43	1.46	-0.03	0.784	0.00	0.972	0.00
Withdrawn Behavior+	0.71	0.83	-0.12	0.075	-0.13*	0.077	-0.13
Total Problem Behavior	4.84	5.05	-0.21	0.450	-0.19	0.453	-0.05

Exhibit C.2A. Estimated Impacts on Social-Emotional Outcomes by Year: 4-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Parent-Reported Measures (cont'd)							
Social Competencies+	11.09	11.13	-0.05	0.534	-0.02	0.753	-0.02
Social Skills and Positive Approaches to Learning	12.64	12.63	0.01	0.931	0.02	0.764	0.02
Closeness	33.21	33.26	-0.04	0.789	-0.01	0.944	0.00
Conflict	16.68	17.20	-0.52	0.368	-0.50	0.373	-0.07
Positive Relationships+	64.42	63.99	0.43	0.498	0.41	0.507	0.05
Teacher-Reported Measures							
ASPI-Aggressive	48.56	49.12	-0.56	0.381	-0.72	0.257	-0.09
ASPI-Inattentive/ Hyperactive	50.35	50.50	-0.15	0.852	-0.26	0.731	-0.03
ASPI-Withdrawn/ Low Energy	49.87	49.22	0.65	0.257	0.75	0.169	0.11
ASPI-Oppositional	47.79	47.88	-0.09	0.909	-0.36	0.637	-0.05
ASPI-Problems with Peer Interaction	51.33	51.53	-0.20	0.804	-0.38	0.630	-0.03
ASPI-Shy/Socially Reticent	48.00	46.76	1.24	0.043	1.37**	0.019	0.19
ASPI-Problems with Structured Learning	51.03	50.29	0.74	0.305	0.74	0.306	0.07
ASPI-Problems with Teacher Interaction	50.14	48.81	1.33	0.106	1.29*	0.099	0.13
Closeness	29.91	29.74	0.17	0.544	0.22	0.465	0.05
Conflict	14.22	13.92	0.30	0.543	0.09	0.838	0.01
Positive Relationships	63.54	63.61	-0.07	0.906	0.20	0.728	0.02
3rd Grade Year (Spring 2007)							
Parent-Reported Measures							
Aggressive Behavior	2.24	2.47	-0.23	0.073	-0.23**	0.043	-0.13
Hyperactive Behavior	1.91	1.99	-0.07	0.520	-0.08	0.435	-0.05
Withdrawn Behavior+	1.02	1.13	-0.11	0.163	-0.11	0.187	-0.09
Total Problem Behavior	5.70	6.18	-0.47	0.137	-0.50*	0.090	-0.12

Exhibit C.2A. Estimated Impacts on Social-Emotional Outcomes by Year: 4-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Social Skills and Positive Approaches to Learning+	11.95	12.11	-0.16	0.208	-0.10	0.383	-0.05
Teacher-Reported Measures							
Conduct Problems-Percent in Normal Category	0.76	0.80	-0.04	0.155	-0.02	0.394	-0.06
Emotional Symptoms-Percent in Normal Category	0.89	0.94	-0.05	0.005	**-0.06*** **	0.005	-0.24
Hyperactivity-Percent in Normal Category	0.76	0.75	0.01	0.758	0.00	0.938	-0.01
Peer Problems-Percent in Normal Category	0.83	0.89	-0.05	0.010	-0.04	0.104	-0.11
Pro-social-Percent in Normal Category	0.74	0.78	-0.04	0.225	-0.05	0.163	-0.13
Total Difficulties-Percent in Normal Category	0.71	0.77	-0.06	0.117	-0.06	0.140	-0.15
Closeness	28.03	28.56	-0.53	0.077	-0.67*	0.060	-0.13
Conflict	14.55	13.72	0.83	0.064	0.65	0.136	0.10
Positive Relationships	61.05	62.41	-1.36	0.034	-1.33*	0.063	-0.14
Social Competency++++	0.02	0.12	-0.10	0.186	-0.09	0.261	-0.09
Child-Reported Measures							
Externalizing++++	0.00	0.09	-0.09	0.249	-0.09	0.226	-0.09
Internalizing++++	0.03	0.14	-0.11	0.199	-0.10	0.212	-0.10
Peer Relations++++	-0.06	0.07	-0.13	0.021	-0.14**	0.020	-0.13
School++++	-0.02	-0.07	0.05	0.586	0.05	0.575	0.05

Key:

*** $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

+ Indicates the reliability (Cronbach alpha) for the measure is < 0.6.

++++ These scores are reported in the Item Response Theory (IRT) metric. Since there is not a normalizing sample to link the scores to, the scores are reported in a standardized format with a mean of 0 and a standard deviation of 1.

Bold regression-adjusted impact value indicates the outcome passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.

Exhibit C.2B. Estimated Impacts on Social-Emotional Outcomes by Year: 3-Year-Old Cohort

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Head Start Baseline (Fall 2002)							
Parent-Reported Measures							
Aggressive Behavior	3.12	3.03					
Hyperactive Behavior	1.86	1.91					
Withdrawn Behavior	0.63	0.60					
Total Problem Behavior	6.16	6.09					
Social Competencies	10.74	10.70					
Social Skills and Positive Approaches to Learning	12.23	12.08					
Head Start Year (Spring 2003)							
Parent-Reported Measures							
Aggressive Behavior	2.97	3.05	-0.08	0.417	-0.10	0.274	-0.06
Hyperactive Behavior	1.71	2.00	-0.29	0.004	**-0.33*** **	0.001	-0.21
Withdrawn Behavior+	0.55	0.58	-0.02	0.708	-0.04	0.510	-0.04
Total Problem Behavior	5.80	6.24	-0.44	0.053	**-0.52*** **	0.003	-0.14
Social Competencies+	10.95	10.99	-0.04	0.540	-0.03	0.637	-0.03
Social Skills and Positive Approaches to Learning	12.41	12.38	0.03	0.740	0.04	0.745	0.02
Closeness	33.63	33.44	0.19	0.178	0.18	0.220	0.06
Conflict	18.04	18.12	-0.07	0.860	-0.05	0.893	-0.01
Positive Relationships+	63.50	63.19	0.31	0.491	0.28	0.517	0.03
Age 4 Year (2004)							
Parent-Reported Measures							
Aggressive Behavior	2.59	2.70	-0.11	0.310	-0.12	0.203	-0.07
Hyperactive Behavior+	1.64	1.73	-0.10	0.349	-0.13	0.242	-0.09
Withdrawn Behavior+	0.59	0.64	-0.05	0.467	-0.08	0.248	-0.08
Total Problem Behavior	5.29	5.63	-0.34	0.197	-0.39	0.115	-0.10

Exhibit C.2B. Estimated Impacts on Social-Emotional Outcomes by Year: 3-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Parent-Reported Measures (cont'd)							
Social Competencies	11.04	11.11	-0.06	0.447	-0.01	0.868	-0.01
Social Skills and Positive Approaches to Learning	12.59	12.46	0.13	0.177	0.19*	0.055	0.11
Closeness	33.53	33.34	0.18	0.304	0.22	0.232	0.08
Conflict	17.67	18.12	-0.44	0.412	-0.39	0.461	-0.06
Positive Relationships+	63.80	63.06	0.73	0.255	0.72	0.261	0.09
Kindergarten Year (Spring 2005)							
Parent-Reported Measures							
Aggressive Behavior	2.32	2.37	-0.05	0.695	-0.08	0.382	-0.05
Hyperactive Behavior	1.37	1.52	-0.14	0.124	-0.18**	0.048	-0.12
Withdrawn Behavior+	0.63	0.65	-0.02	0.693	-0.03	0.563	-0.03
Total Problem Behavior	4.84	5.06	-0.22	0.405	-0.26	0.246	-0.07
Social Competencies+	11.06	10.96	0.10	0.209	0.11	0.179	0.08
Social Skills and Positive Approaches to Learning	12.55	12.29	0.26	0.076	0.25*	0.075	0.14
Closeness	33.19	33.05	0.14	0.390	0.13	0.434	0.05
Conflict	17.14	17.13	0.01	0.987	-0.06	0.888	-0.01
Positive Relationships	63.88	63.82	0.06	0.893	0.12	0.803	0.02
Teacher-Reported Measures							
ASPI-Aggressive	49.02	48.66	0.36	0.444	0.40	0.318	0.05
ASPI-Inattentive/ Hyperactive	50.27	50.37	-0.10	0.852	-0.02	0.972	0.00
ASPI-Withdrawn/ Low Energy	49.09	48.66	0.44	0.320	0.52	0.241	0.08
ASPI-Oppositional	48.30	48.36	-0.05	0.896	0.03	0.953	0.00
ASPI-Problems with Peer Interaction	51.49	50.94	0.55	0.449	0.64	0.335	0.06
ASPI-Shy/Socially Reticent	47.37	47.44	-0.07	0.888	0.07	0.882	0.01
ASPI-Problems with Structured Learning	50.40	49.83	0.57	0.309	0.74	0.136	0.07

Exhibit C.2B. Estimated Impacts on Social-Emotional Outcomes by Year: 3-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Teacher-Reported Measures (cont'd)							
ASPI-Problems with Teacher Interaction	49.42	49.28	0.14	0.833	0.27	0.649	0.03
Closeness	30.02	30.25	-0.23	0.401	-0.26	0.295	-0.06
Conflict	13.99	14.02	-0.03	0.961	0.03	0.959	0.00
Positive Relationships	63.82	64.14	-0.32	0.613	-0.40	0.483	-0.04
1st Grade Year (Spring 2006)							
Parent-Report Measures							
Aggressive Behavior	2.24	2.27	-0.03	0.775	-0.05	0.624	-0.03
Hyperactive Behavior	1.38	1.49	-0.10	0.219	-0.11	0.127	-0.07
Withdrawn Behavior+	0.73	0.71	0.02	0.641	0.02	0.732	0.02
Total Problem Behavior	4.88	5.01	-0.13	0.577	-0.15	0.439	-0.04
Social Competencies	11.13	11.09	0.05	0.559	0.08	0.317	0.07
Social Skills and Positive Approaches to Learning	12.58	12.51	0.06	0.564	0.05	0.642	0.03
Closeness	33.32	33.09	0.23	0.066	0.29**	0.013	0.10
Conflict	16.75	17.28	-0.53	0.268	-0.55	0.210	-0.08
Positive Relationships	64.46	63.77	0.68	0.172	0.77*	0.098	0.10
Teacher-Reported Measures							
ASPI-Aggressive	48.81	49.15	-0.34	0.528	-0.54	0.266	-0.07
ASPI-Inattentive/ Hyperactive	50.38	50.67	-0.29	0.597	-0.45	0.402	-0.05
ASPI-Withdrawn/ Low Energy	49.51	48.95	0.57	0.287	0.44	0.383	0.06
ASPI-Oppositional	48.39	48.25	0.14	0.835	0.04	0.944	0.01
ASPI-Problems with Peer Interactions	51.96	52.20	-0.24	0.792	-0.43	0.584	-0.04
ASPI-Shy/Socially Reticent	47.36	47.05	0.31	0.533	0.21	0.623	0.03
ASPI-Problems with Structured Learning	50.76	50.55	0.21	0.789	-0.12	0.855	-0.01
ASPI-Problems with Teacher Interaction	50.00	50.13	-0.12	0.847	-0.15	0.819	-0.01

Exhibit C.2B. Estimated Impacts on Social-Emotional Outcomes by Year: 3-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Teacher-Reported Measures (cont'd)							
Closeness	29.94	29.74	0.20	0.475	0.28	0.301	0.06
Conflict	14.12	14.13	-0.01	0.983	-0.16	0.722	-0.02
Positive Relationships	63.56	63.37	0.19	0.780	0.44	0.461	0.05
3rd Grade Year (Spring 2008)							
Parent-Report Measures							
Aggressive Behavior	2.42	2.38	0.04	0.752	0.04	0.703	0.02
Hyperactive Behavior	1.88	1.87	0.01	0.931	0.01	0.942	0.00
Withdrawn Behavior+	1.05	0.93	0.12	0.151	0.10	0.230	0.08
Total Problem Behavior	5.94	5.77	0.17	0.600	0.15	0.604	0.03
Social Skills and Positive Approaches to Learning+	12.19	11.97	0.23	0.069	0.24**	0.025	0.12
Teacher-Reported Measures							
Conduct Problems-Percent in Normal Category	0.76	0.75	0.02	0.629	-0.01	0.759	-0.02
Emotional Symptoms-Percent in Normal Category	0.92	0.89	0.03	0.122	0.01	0.705	0.03
Hyperactivity-Percent in Normal Category	0.77	0.73	0.04	0.207	0.02	0.431	0.05
Peer Problems-Percent in Normal Category	0.85	0.82	0.03	0.128	0.00	0.984	0.00
Pro-social-Percent in Normal Category	0.74	0.75	-0.01	0.688	-0.05	0.137	-0.12
Total Difficulties Percent in Normal Category	0.73	0.70	0.03	0.279	0.01	0.832	0.01
Closeness	28.08	28.13	-0.05	0.899	-0.40	0.300	-0.08
Conflict	14.48	14.62	-0.14	0.774	0.27	0.613	0.04
Positive Relationships	61.34	61.09	0.25	0.712	-0.45	0.549	-0.04

Exhibit C.2B. Estimated Impacts on Social-Emotional Outcomes by Year: 3-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Social Competency++++	0.02	-0.00	0.02	0.737	-0.05	0.427	-0.05
Child-Reported Measures							
Externalizing++++	-0.04	-0.02	-0.02	0.763	-0.02	0.733	-0.02
Internalizing++++	-0.05	-0.07	0.02	0.753	0.02	0.731	0.02
Peer Relations++++	0.04	-0.05	0.08	0.230	0.08	0.227	0.09
School++++	0.09	0.05	0.04	0.596	0.05	0.564	0.04

Key:

******* $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

+ Indicates the reliability (Cronbach alpha) for the measure is < 0.6.

++++ These scores are reported in the Item Response Theory (IRT) metric. Since there is not a normalizing sample to link the scores to, the scores are reported in a standardized format with a mean of 0 and a standard deviation of 1.

Bold regression-adjusted impact value indicates the outcome passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.

Exhibit C.3A. Estimated Impacts on Parent-Reported Child Health Outcomes by Year: 4-Year-Old Cohort

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Head Start Baseline (Fall 2002)							
Parent-Reported Measures							
Child Received Dental Care	0.83	0.67					
Child Has Health Insurance Coverage	0.84	0.89					
Child's Overall Health Status Is Excellent/Good	0.79	0.79					
Child Needs Ongoing Care	0.13	0.19					
Child Had Care for Injury Last Month	0.06	0.05					
Head Start Year (Spring 2003)							
Parent-Reported Measures							
Child Received Dental Care	0.73	0.56	0.17	0.000	**0.15*****	0.000	0.31
Child Has Health Insurance Coverage	0.89	0.88	0.01	0.473	0.01	0.733	0.02
Child's Overall Health Status Is Excellent/Good	0.79	0.82	-0.03	0.225	-0.03	0.244	-0.07
Child Needs Ongoing Care	0.12	0.11	0.01	0.656	0.01	0.422	0.05
Child Had Care for Injury Last Month	0.12	0.12	0.00	0.929	-0.02	0.409	-0.06
Kindergarten Year (Spring 2004)							
Parent-Reported Measures							
Child Received Dental Care	0.67	0.66	0.01	0.670	0.03	0.435	0.06
Child Has Health Insurance Coverage	0.90	0.86	0.04	0.060	0.04*	0.056	0.11

Exhibit C.3A. Estimated Impacts on Parent-Reported Child Health Outcomes by Year: 4-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Parent-Reported Measures (cont'd)							
Child's Overall Health Status Is Excellent/Good	0.82	0.76	0.06	0.094	0.05*	0.098	0.13
Child Needs Ongoing Care	0.12	0.13	-0.01	0.619	-0.02	0.432	-0.06
Child Had Care for Injury Last Month	0.12	0.10	0.02	0.380	0.02	0.547	0.05
1st Grade Year (Spring 2005)							
Parent-Reported Measures							
Child Received Dental Care	0.68	0.66	0.02	0.524	0.02	0.550	0.03
Child Has Health Insurance Coverage	0.89	0.85	0.04	0.073	0.04**	0.044	0.11
Child's Overall Health Status Is Excellent/Good	0.81	0.81	0.00	0.951	-0.01	0.858	-0.01
Child Needs Ongoing Care	0.14	0.13	0.02	0.476	0.02	0.291	0.07
Child Had Care for Injury Last Month	0.15	0.13	0.02	0.260	0.02	0.303	0.06
3rd Grade Year (Spring 2007)							
Parent-Reported Measures							
Child Received Dental Care	0.86	0.85	0.01	0.674	0.01	0.730	0.03
Child Has Health Insurance Coverage	0.88	0.87	0.01	0.648	0.00	1.000	0.00
Child's Overall Health Status Is Excellent/Good	0.80	0.79	0.02	0.478	0.01	0.547	0.03

Exhibit C.3A. Estimated Impacts on Parent-Reported Child Health Outcomes by Year: 4-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Child Needs Ongoing Care	0.16	0.15	0.01	0.650	0.00	0.861	0.01
Child Had Care for Injury Last Month	0.12	0.11	0.01	0.724	0.01	0.815	0.02

Key:

*** $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

Bold regression-adjusted impact value indicates the outcome passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.

Exhibit C.3B. Estimated Impacts on Child Health Outcomes by Year: 3-Year-Old Cohort

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Head Start Baseline (Fall 2002)							
Parent-Reported Measures							
Child Received Dental Care	0.76	0.58					
Child Has Health Insurance Coverage	0.91	0.87					
Child's Overall Health Status Is Excellent/Good	0.80	0.78					
Child Needs Ongoing Care	0.18	0.22					
Child Had Care for Injury Last Month	0.06	0.07					
Head Start Year (Spring 2003)							
Parent-Reported Measures							
Child Received Dental Care	0.68	0.52	0.17	0.000	**0.17*** **	0.000	0.33
Child Has Health Insurance Coverage	0.92	0.92	0.01	0.738	0.00	0.803	0.01
Child's Overall Health Status Is Excellent/Good	0.81	0.76	0.05	0.084	0.05**	0.045	0.11
Child Needs Ongoing Care	0.13	0.13	0.00	0.910	0.00	0.988	0.00
Child Had Care for Injury Last Month	0.09	0.09	0.00	0.918	-0.01	0.699	-0.02
Age 4 Year (Spring 2004)							
Parent-Reported Measures							
Child Received Dental Care	0.74	0.65	0.09	0.001	**0.10*** **	0.001	0.20
Child Has Health Insurance Coverage	0.93	0.92	0.01	0.460	0.00	0.935	0.00
Child's Overall Health Status Is Excellent/Good	0.83	0.83	0.00	0.897	0.00	0.851	0.01
Child Needs Ongoing Care	0.13	0.13	0.00	0.862	0.01	0.739	0.02

Exhibit C.3B. Estimated Impacts on Child Health Outcomes by Year: 3-Year-Old Cohort (continued)

	Mean Estimates				Regression-Adjusted Impact		
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	Effect Size
Parent-Reported Measures (cont'd)							
Child Had Care for Injury Last Month	0.12	0.09	0.03	0.039	0.03*	0.089	0.10
Kindergarten Year (Spring 2005)							
Parent-Reported Measures							
Child Received Dental Care	0.75	0.72	0.04	0.114	0.03	0.270	0.06
Child Has Health Insurance Coverage	0.93	0.90	0.03	0.087	0.04**	0.044	0.14
Child's Overall Health Status Is Excellent/Good	0.81	0.83	-0.01	0.660	0.00	0.889	-0.01
Child Needs Ongoing Care	0.15	0.19	-0.04	0.051	-0.03	0.114	-0.07
Child Had Care for Injury Last Month	0.13	0.13	0.00	0.836	0.00	0.985	0.00
1st Grade Year (Spring 2006)							
Parent-Reported Measures							
Child Received Dental Care	0.74	0.73	0.02	0.514	0.01	0.786	0.02
Child Has Health Insurance Coverage	0.93	0.92	0.02	0.187	0.02	0.252	0.06
Child's Overall Health Status Is Excellent/Good	0.84	0.84	0.00	0.900	0.02	0.434	0.04
Child Needs Ongoing Care	0.16	0.17	-0.01	0.705	-0.01	0.578	-0.03
Child Had Care for Injury Last Month	0.09	0.07	0.02	0.370	0.02	0.294	0.07
3rd Grade Year (Spring 2008)							
Parent-Reported Measures							
Child Received Dental Care	0.87	0.86	0.01	0.681	0.00	0.979	0.00
Child Has Health Insurance Coverage	0.90	0.89	0.01	0.476	0.02	0.354	0.06

Exhibit C.3B. Estimated Impacts on Child Health Outcomes by Year: 3-Year-Old Cohort (continued)

	Mean Estimates				Regression-Adjusted Impact		
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	Effect Size
Child's Overall Health Status Is Excellent/Good	0.80	0.81	-0.01	0.714	-0.01	0.681	-0.02
Child Needs Ongoing Care	0.19	0.15	0.03	0.179	0.02	0.421	0.06
Child Had Care for Injury Last Month	0.12	0.16	-0.04	0.217	-0.04	0.216	-0.11

Key:

*** $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

Bold regression-adjusted impact value indicates the outcome passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.

Exhibit C.4A. Estimated Impacts on Parenting Outcomes by Year: 4-Year-Old Cohort

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Head Start Baseline (Fall 2002)							
Parent-Reported Measures							
Parent Spanked Child in Last Week	0.41	0.43					
Parent Used Time Out in Last Week	0.62	0.61					
Parent Read to Child in Last Week	0.37	0.34					
Parental Safety Practices Scale	3.66	3.66					
Family Cultural Enrichment Scale	3.44	3.41					
Head Start Year (Spring 2003)							
Parent-Reported Measures							
Parent Spanked Child in Last Week	0.38	0.38	0.00	0.946	-0.01	0.750	-0.02
Parent Used Time Out in Last Week	0.61	0.68	-0.07	0.044	-0.08**	0.025	-0.17
Parent Read to Child in Last Week	0.36	0.33	0.03	0.391	0.03	0.396	0.06
Parental Safety Practices Scale+	3.73	3.71	0.02	0.364	0.03	0.382	0.08
Family Cultural Enrichment Scale+	4.01	3.89	0.11	0.237	0.08	0.368	0.06
Kindergarten Year (Spring 2004)							
Parent-Reported Measures							
Parent Spanked Child in Last Week	0.29	0.31	-0.02	0.481	0.00	0.869	-0.01
Parent Used Time Out in Last Week	0.58	0.57	0.01	0.858	0.01	0.689	0.02
Parent Read to Child in Last Week	0.35	0.38	-0.03	0.440	-0.03	0.385	-0.07
Parental Safety Practices Scale+	3.72	3.68	0.03	0.251	0.04	0.156	0.11
Family Cultural Enrichment Scale+	4.11	3.97	0.14	0.153	0.14	0.142	0.10
Parenting Style: Authoritarian	0.08	0.08	0.00	0.873	0.00	0.981	0.00
Parenting Style: Authoritative	0.68	0.63	0.05	0.205	0.05	0.164	0.10

Exhibit C.4A. Estimated Impacts on Parenting Outcomes by Year: 4-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Parenting Style: Neglectful	0.09	0.12	-0.03	0.200	-0.03	0.211	-0.09
Parenting Style: Permissive	0.18	0.19	-0.01	0.548	-0.02	0.447	-0.05
Teacher-Reported Measures							
School Contact and Communication	0.83	0.84	-0.01	0.848	-0.01	0.845	-0.02
Parent Participation	0.89	0.90	-0.01	0.557	-0.01	0.841	-0.02
1st Grade Year (Spring 2005)							
Parent-Reported Measures							
Parent Spanked Child in Last Week	0.21	0.22	0.00	0.881	0.00	0.976	0.00
Parent Used Time Out in Last Week	0.49	0.53	-0.04	0.305	-0.04	0.322	-0.08
Parent Read to Child in Last Week	0.40	0.43	-0.03	0.407	-0.01	0.733	-0.02
Family Cultural Enrichment Scale+	4.00	3.94	0.06	0.529	0.04	0.612	0.03
Parenting Style: Authoritarian	0.07	0.10	-0.03	0.156	-0.03	0.199	-0.10
Parenting Style: Authoritative	0.67	0.64	0.03	0.302	0.04	0.158	0.09
Parenting Style: Neglectful	0.07	0.09	-0.01	0.416	-0.02	0.327	-0.06
Parenting Style: Permissive	0.19	0.18	0.01	0.579	0.00	0.936	0.00
Teacher-Reported Measures							
School Contact and Communication	0.79	0.81	-0.02	0.635	-0.02	0.570	-0.06
Parent Participation	0.86	0.87	-0.01	0.753	-0.01	0.817	-0.02
3rd Grade Year (Spring 2007)							
Parent-Reported Measures							
Parent Spanked Child in Last Week	0.23	0.25	-0.02	0.595	-0.02	0.635	-0.04
Parenting Style: Authoritarian	0.16	0.13	0.02	0.333	0.02	0.494	0.05
Parenting Style: Authoritative	0.56	0.60	-0.04	0.340	-0.02	0.516	-0.05

Exhibit C.4A. Estimated Impacts on Parenting Outcomes by Year: 4-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Parenting Style: Neglectful	0.12	0.12	0.00	0.960	0.00	0.843	-0.01
Parenting Style: Permissive	0.16	0.14	0.02	0.495	0.01	0.523	0.04
Supportive School Environment++++	-0.00	-0.03	0.03	0.620	0.03	0.701	0.03
Effect of Parenting on Parent's Life++++	0.05	0.01	0.04	0.603	0.06	0.450	0.06
Doing Things Together ++++	-0.06	-0.04	-0.02	0.796	-0.02	0.786	-0.02
Time Spent with Child++++	0.05	-0.19	0.24	0.003	**0.27*** **	0.001	0.27
Parent Perception of School Services++++	-0.10	-0.02	-0.09	0.145	-0.09	0.175	-0.10
Teacher-Reported Measures							
School Contact and Communication	0.66	0.72	-0.05	0.058	-0.04	0.103	-0.10
Parent Participation	0.81	0.86	-0.05	0.097	-0.01	0.652	-0.04

Key:

*** $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

+ Indicates the reliability (Cronbach alpha) for the measure is <0.6.

++++ These scores are reported in the Item Response Theory (IRT) metric. Since there is not a normalizing sample to link the scores to, the scores are reported in a standardized format with a mean of 0 and a standard deviation of 1.

Bold regression-adjusted impact value indicates the outcome passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.

Exhibit C.4B. Estimated Impacts on Parenting Outcomes by Year: 3-Year-Old Cohort

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Head Star Baseline (Fall 2002)							
Parent-Reported Measures							
Parent Spanked Child in Last Week	0.45	0.50					
Parent Used Time Out in Last Week	0.61	0.66					
Parent Read to Child in Last Week	0.36	0.32					
Parental Safety Practices Scale	3.67	3.65					
Family Cultural Enrichment Scale	3.41	3.24					
Head Start Year (Spring 2003)							
Parent-Reported Measures							
Parent Spanked Child in Last Week	0.42	0.48	-0.07	0.037	**-0.07****	0.025	-0.14
Parent Used Time Out in Last Week	0.62	0.66	-0.04	0.211	-0.04	0.205	-0.08
Parent Read to Child in Last Week	0.35	0.28	0.07	0.045	**0.07****	0.030	0.15
Parental Safety Practices Scale+	3.73	3.70	0.03	0.170	0.03	0.146	0.10
Family Cultural Enrichment Scale+	3.78	3.55	0.23	0.003	**0.25*****	0.000	0.18
Age 4 Year (Spring 2004)							
Parent-Reported Measures							
Parent Spanked Child in Last Week	0.35	0.33	0.02	0.602	0.01	0.635	0.03
Parent Used Time Out in Last Week	0.61	0.64	-0.03	0.338	-0.02	0.355	-0.05
Parent Read to Child in Last Week	0.34	0.35	0.00	0.917	-0.01	0.827	-0.01
Parental Safety Practices Scale+	3.73	3.71	0.03	0.282	0.02	0.313	0.06
Family Cultural Enrichment Scale+	3.94	3.87	0.07	0.415	0.04	0.593	0.03
Parenting Style: Authoritarian	0.04	0.08	-0.04	0.006	**-0.04*****	0.005	-0.14
Parenting Style: Authoritative	0.72	0.67	0.05	0.081	0.04	0.186	0.08

Exhibit C.4B. Estimated Impacts on Parenting Outcomes by Year: 3-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Parent-Reported Measures (cont'd)							
Parenting Style: Neglectful	0.06	0.06	0.00	0.823	0.00	0.826	-0.01
Parenting Style: Permissive	0.19	0.20	0.00	0.882	0.00	0.893	0.01
Kindergarten Year (Spring 2005)							
Parent-Reported Measures							
Parent Spanked Child in Last Week	0.26	0.31	-0.04	0.073	-0.04*	0.070	-0.09
Parent Used Time Out in Last Week	0.52	0.58	-0.06	0.022	-0.07**	0.013	-0.13
Parent Read to Child in Last Week	0.36	0.32	0.03	0.302	0.03	0.305	0.07
Parental Safety Practices Scale+	3.72	3.71	0.01	0.673	0.01	0.714	0.02
Family Cultural Enrichment Scale+	3.93	3.93	0.00	0.988	0.00	0.968	0.00
Parenting Style: Authoritarian	0.06	0.07	-0.01	0.545	0.00	0.950	0.00
Parenting Style: Authoritative	0.67	0.66	0.01	0.656	0.00	0.905	-0.01
Parenting Style: Neglectful	0.06	0.08	-0.02	0.229	-0.02	0.202	-0.09
Parenting Style: Permissive	0.21	0.19	0.02	0.475	0.03	0.310	0.07
Teacher-Reported Measures							
School Contact and Communication	0.82	0.82	0.00	0.957	0.00	0.879	0.01
Parent Participation+	0.87	0.87	-0.01	0.806	0.00	0.886	-0.01
1st Grade Year (Spring 2006)							
Parent-Reported Measures							
Parent Spanked Child in Last Week	0.21	0.23	-0.03	0.252	-0.03	0.183	-0.07
Parent Used Time Out in Last Week	0.48	0.53	-0.06	0.080	-0.05*	0.075	-0.11
Parent Read to Child in Last Week	0.38	0.36	0.02	0.525	0.01	0.746	0.02
Family Cultural Enrichment Scale+	3.92	3.87	0.05	0.590	0.01	0.879	0.01

Exhibit C.4B. Estimated Impacts on Parenting Outcomes by Year: 3-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Parent-Reported Measures (cont'd)							
Parenting Style: Authoritarian	0.05	0.07	-0.03	0.031	-0.03**	0.046	-0.11
Parenting Style: Authoritative	0.71	0.69	0.03	0.389	0.00	0.877	0.01
Parenting Style: Neglectful	0.06	0.06	0.00	0.758	0.00	0.974	0.00
Parenting Style: Permissive	0.18	0.17	0.01	0.743	0.02	0.320	0.06
Teacher-Reported Measures							
School Contact and Communication	0.82	0.79	0.03	0.353	0.02	0.453	0.05
Parent Participation+	0.85	0.84	0.01	0.739	0.01	0.626	0.04
3rd Grade Year (Spring 2008)							
Parent-Reported Measures							
Parent Spanked Child in Last Week	0.25	0.28	-0.02	0.467	-0.02	0.630	-0.04
Parenting Style: Authoritarian	0.11	0.14	-0.03	0.307	-0.03	0.298	-0.08
Parenting Style: Authoritative	0.67	0.59	0.08	0.035	0.08**	0.033	0.16
Parenting Style: Neglectful	0.06	0.08	-0.02	0.403	-0.01	0.495	-0.05
Parenting Style: Permissive	0.15	0.19	-0.04	0.172	-0.04	0.160	-0.10
Supportive School Environment++++	0.05	-0.00	0.05	0.378	0.03	0.672	0.03
Effect of Parenting on Parent's Life++++	-0.06	-0.03	-0.03	0.531	-0.04	0.429	-0.04
Doing Things Together ++++	0.00	-0.07	0.08	0.208	0.08	0.160	0.09
Time Spent with Child++++	-0.04	-0.01	-0.03	0.643	-0.05	0.408	-0.05
Parent Perception of School Services++++	0.09	-0.03	0.12	0.084	0.11	0.142	0.11

Exhibit C.4B. Estimated Impacts on Parenting Outcomes by Year: 3-Year-Old Cohort (continued)

Outcomes	Mean Estimates				Regression-Adjusted Impact		Effect Size
	Head Start Group	Control Group	Head Start - Control	p-value	Impact	p-value	
Teacher-Reported Measures							
School Contact and Communication	0.77	0.75	0.02	0.501	0.00	0.883	-0.01
Parent Participation+	0.81	0.84	-0.03	0.203	-0.03	0.181	-0.08

Key:

******* $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

+ Indicates the reliability (Cronbach alpha) for the measure is <0.06.

++++ These scores are reported in the Item Response Theory (IRT) metric. Since there is not a normalizing sample to link the scores to, the scores are reported in a standardized format with a mean of 0 and a standard deviation of 1.

Bold regression-adjusted impact value indicates the outcome passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.

Appendix D

Impact on the Treated (IOT) Tables, 2003-2008

Exhibit D.1A. Estimated IOT Impacts on Cognitive Outcomes, by Year: 4-Year-Old Cohort

Outcome	ITT Regression Adjusted Impact	IOT		
	Impact	Impact	p-value	Effect Size
Head Start Year (Spring 2003)				
Language and Literacy Measures				
PPVT (Adapted)	3.55	**5.31****	0.028	0.13
WJ-III Letter-Word Identification	5.98	**8.94****	0.017	0.34
WJ-III Spelling	3.77	**5.64****	0.029	0.22
WJ-III Oral Comprehension	-0.94	-1.41	0.395	-0.08
CTOPPP Elision	2.45	3.66	0.444	0.07
Color Identification	0.08	**0.12*****	0.010	0.25
Letter Naming	2.36	**3.53*****	0.002	0.37
WJ-III Pre-Academic Skills	4.23	**6.32****	0.022	0.29
Spanish Language and Literacy Measures				
TVIP (Adapted)	9.04	10.65	0.106	0.25
WM Letter-Word Identification	1.91	2.25	0.180	0.16
Math Skills Measures				
WJ-III Applied Problems	3.18	4.75	0.139	0.18
Counting Bears	0.04	0.06	0.181	0.12
Fine Motor Skills Measures				
McCarthy Draw-a-Design	0.20	0.30	0.110	0.15
Parent-Reported Literacy Measure				
Emergent Literacy Scale	0.43	**0.64*****	0.000	0.46
Kindergarten Year (Spring 2004)				
Language and Literacy Measures				
PPVT (Adapted)	1.78	2.66	0.328	0.06
WJ-III Letter-Word Identification	-0.19	-0.28	0.918	-0.01
WJ-III Spelling	-0.52	-0.78	0.764	-0.03
WJ-III Oral Comprehension	-0.91	-1.36	0.327	-0.08
CTOPPP Elision	-2.85	-4.26	0.374	-0.09
Letter Naming	0.40	0.60	0.274	0.09
WJ-III Pre-Academic Skills	-0.47	-0.70	0.745	-0.03
WJ-III Word Attack	-1.13	-1.69	0.639	-0.05
WJ-III Basic Reading Skills	-0.71	-1.06	0.728	-0.03
Spanish Language and Literacy Measures				
TVIP (Adapted)	-1.03	-1.21	0.868	-0.02
WM Letter-Word Identification	-4.28	-5.04	0.130	-0.19

Exhibit D.1A. Estimated IOT Impacts on Cognitive Outcomes, by Year: 4-Year-Old Cohort (continued)

Outcome	ITT Regression Adjusted Impact	IOT		
	Impact	Impact	p-value	Effect Size
Math Skills Measures				
WJ-III Applied Problems	0.12	0.18	0.936	0.01
WJ-III Quantitative Concepts	-0.13	-0.19	0.920	-0.01
WJ-III Math Reasoning	-0.08	-0.12	0.951	-0.01
School Performance Assessment Measures				
School Accomplishments	0.00	0.00	0.997	0.00
Promotion	0.00	0.00	0.888	0.00
Language and Literacy Ability	0.04	0.06	0.424	0.13
Math Ability	0.05	0.07	0.191	0.17
Social Studies and Science Ability	0.03	0.07	0.501	0.19
1st Grade Year (Spring 2005)				
Language and Literacy Measures				
PPVT (Adapted)	2.95	4.41*	0.072	0.14
WJ-III Letter-Word Identification	0.83	1.24	0.705	0.03
WJ-III Spelling	1.55	2.32	0.347	0.09
WJ-III Oral Comprehension	0.34	0.51	0.717	0.03
WJ-III Pre-Academic Skills	0.95	1.42	0.510	0.06
WJ-III Word Attack	1.71	2.56	0.324	0.08
WJ-III Basic Reading Skills	1.08	1.61	0.550	0.05
WJ-III Academic Applications	0.38	0.57	0.730	0.03
WJ-III Academic Skills	1.11	1.66	0.446	0.07
WJ-III Passage Comprehension	0.17	0.25	0.922	0.01
WJ-III Writing Sample	0.15	0.22	0.824	0.02
Spanish Language and Literacy Measures				
TVIP (Adapted)	5.25	6.18	0.240	0.15
WM Letter-Word Identification	-4.30	-5.06	0.397	-0.11
Math Skills Measures				
WJ-III Applied Problems	0.82	1.23	0.523	0.06
WJ-III Quantitative Concepts	0.32	0.48	0.819	0.03
WJ-III Math Reasoning	0.47	0.70	0.705	0.04
WJ-III Calculation	1.41	2.11	0.255	0.11

Exhibit D.1A. Estimated IOT Impacts on Cognitive Outcomes, by Year: 4-Year-Old Cohort (continued)

	ITT Regression Adjusted Impact	IOT		
Outcome	**Impact**	**Impact**	**p-value**	**Effect Size**
School Performance Assessment Measures				
School Accomplishments	-0.59	-0.88	0.500	-0.09
Promotion	0.01	0.01	0.376	0.06
Language and Literacy Ability	-0.02	-0.03	0.433	-0.07
Math Ability	-0.05	-0.07	0.148	-0.19
Social Studies and Science Ability	-0.02	-0.03	0.362	-0.08
3rd Grade Year (Spring 2007)				
Language and Literacy Measures				
ECLS-K Reading	2.23	3.34*	0.075	0.16
PPVT (Adapted)	2.17	3.25	0.246	0.12
WJ-III Letter-Word Identification	2.11	3.16	0.275	0.10
Spanish Language and Literacy Measures0.10				
WM Letter-Word Identification	3.53	5.28	0.678	0.10
Math Skills Measures				
WJ-III Applied Problems	-0.43	-0.64	0.729	-0.03
WJ-III Calculation	0.00	0.00	0.997	0.00
School Performance Assessment Measures				
Promotion	0.01	0.01	0.768	0.03
Language and Literacy Ability	0.04	0.06	0.389	0.13
Math Ability	-0.03	-0.04	0.462	-0.10
Reading/Language Arts Skills++	-0.01	-0.01	0.945	-0.01
Math Skills++	-0.03	-0.04	0.632	-0.04

Note: The four-year old cohort compliance rate is 0.66897.

Key:

******* $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

++These scores are reported in the Item Response Theory (IRT) metric. Since there is not a normalizing sample to link the scores to, the scores are reported in a standardized format with a mean of 0 and a standard deviation of 1.

Bold IOT impact indicates the outcome passes the Benjamini-Hochberg test for multiple comparisons with a 10-percent false discovery rate.

Exhibit D.1B. Estimated IOT Impacts on Cognitive Outcomes, by Year: 3-Year-Old Cohort

Outcome	ITT Regression Adjusted Impact	IOT		
	Impact	Impact	p-value	Effect Size
Head Start Year (Spring 2003)				
Language and Literacy Measures				
PPVT (Adapted)	6.53	**9.36*** **	0.000	0.25
WJ-III Letter-Word Identification	6.14	**8.80*** **	0.000	0.37
WJ-III Spelling	2.28	3.05	0.130	0.13
WJ-III Oral Comprehension	0.28	0.40	0.698	0.03
CTOPPP Elision	5.01	**7.18***	0.061	0.14
Color Identification	0.04	0.06	0.179	0.12
Letter Naming	1.56	**2.24*** **	0.005	0.34
WJ-III Pre-Academic Skills	4.25	**6.09*** **	0.004	0.31
Spanish Language and Literacy Measures				
TVIP (Adapted)	5.21	6.09	0.365	0.15
WM Letter-Word Identification	1.59	1.86	0.380	0.15
Pre-Writing Measure				
McCarthy Draw-a-Design	0.16	**0.22*** **	0.007	0.19
Parent-Reported Literacy Measure				
Emergent Literacy Scale	0.48	**0.69*** **	0.000	0.50
Math Skills Measures				
WJ-III Applied Problems	4.35	**6.24****	0.012	0.21
Counting Bears	0.03	0.04	0.241	0.10
Age 4 Year (Spring 2004)				
Language and Literacy Measures				
PPVT (Adapted)	2.03	2.91	0.251	0.07
WJ-III Letter-Word Identification	2.56	3.67	0.112	0.13
WJ-III Spelling	0.28	0.40	0.875	0.02
WJ-III Oral Comprehension	0.25	0.36	0.743	0.02
CTOPPP Elision	8.26	**11.84*** **	0.002	0.21
Color Identification	0.01	0.01	0.466	0.04
Letter Naming	0.85	1.22	0.155	0.13
WJ-III Pre-Academic Skills	1.24	1.78	0.378	0.08
Spanish Language and Literacy Measures				
TVIP (Adapted)	-1.33	-1.55	0.803	-0.03
WM Letter-Word Identification	3.05	3.56	0.334	0.19
Pre-Writing Measure				
McCarthy Draw-a-Design	-0.09	-0.13	0.482	-0.06

Exhibit D.1B. Estimated IOT Impacts on Cognitive Outcomes, by Year: 3-Year-Old Cohort (continued)

Outcome	ITT Regression Adjusted Impact	IOT		
	Impact	Impact	p-value	Effect Size
Parent-Reported Literacy Measure				
Emergent Literacy Scale	0.20	**0.29*** **	0.002	0.22
Math Skills Measures				
WJ-III Applied Problems	0.75	1.08	0.551	0.05
Counting Bears	0.01	0.01	0.777	0.03
Kindergarten Year (Spring 2005)				
Language and Literacy Measures				
PPVT (Adapted)	0.26	0.37	0.851	0.01
WJ-III Letter-Word Identification	0.24	0.34	0.899	0.01
WJ-III Spelling	0.45	0.65	0.774	0.03
WJ-III Oral Comprehension	0.50	0.72	0.633	0.04
CTOPPP Elision	-3.52	-5.05	0.241	-0.11
Letter Naming	-0.32	-0.46	0.340	-0.08
WJ-III Pre-Academic Skills	-0.02	-0.03	0.988	-0.00
WJ-III Word Attack	-1.37	-1.96	0.563	-0.06
WJ-III Basic Reading Skills Composite	-0.54	-0.77	0.801	-0.03
Spanish Language and Literacy Measures				
TVIP (Adapted)	-7.51	-8.78	0.117	-0.22
WM Letter-Word Identification	8.73*	10.20*	0.053	0.30
Math Skills Measures				
WJ-III Applied Problems	-0.94	-1.35	0.519	-0.06
WJ-III Quantitative Concepts	-0.88	-1.26	0.310	-0.08
WJ-III Math Reasoning	-0.91	-1.30	0.408	-0.08
School Performance Assessment Measures				
School Accomplishments	-0.65	-0.93	0.203	-0.13
Promotion	-0.01	-0.01	0.709	-0.05
Language and Literacy Ability	-0.04	-0.06	0.127	-0.14
Math Ability	-0.07	**-0.10*** **	0.003	-0.29
Social Studies and Science Ability	-0.03	-0.04	0.121	-0.13
1st Grade Year (Spring 2006)				
Language and Literacy Measures				
PPVT (Adapted)	2.32	3.33	0.151	0.11
WJ-III Letter-Word Identification	0.37	0.53	0.848	0.02
WJ-III Spelling	-1.20	-1.72	0.438	-0.07
WJ-III Oral Comprehension	1.35	1.94*	0.051	0.12

Exhibit D.1B. Estimated IOT Impacts on Cognitive Outcomes, by Year: 3-Year-Old Cohort (continued)

Outcome	ITT Regression Adjusted Impact	IOT		
	Impact	Impact	p-value	Effect Size
Language and Literacy Measures (cont'd)				
WJ-III Pre-Academic Skills	0.24	0.34	0.869	0.01
WJ-III Word Attack	-0.60	-0.86	0.759	-0.03
WJ-III Basic Reading Skills	-0.08	-0.11	0.966	-0.00
WJ-III Academic Applications Composite	0.73	1.05	0.489	0.06
WJ-III Academic Skills Composite	-0.60	-0.86	0.633	-0.04
WJ-III Passage Comprehension	0.76	1.09	0.580	0.05
WJ-III Writing Sample+	-0.09	-0.13	0.928	-0.01
Spanish Language and Literacy Measures				
TVIP (Adapted)	0.04	0.05	0.993	0.00
WM Letter-Word Identification	-0.54	-0.63	0.910	-0.01
Math Skills Measures				
WJ-III Applied Problems	1.58	2.27	0.163	0.11
WJ-III Quantitative Concepts	0.78	1.12	0.450	0.06
WJ-III Math Reasoning	1.20	1.72	0.231	0.10
WJ-III Calculations	-0.03	-0.04	0.977	-0.00
School Performance Assessment Measures				
School Accomplishments	-0.29	-0.42	0.705	-0.04
Promotion	-0.02	-0.03	0.248	-0.12
Language and Literacy Ability	0.00	0.00	0.888	0.00
Math Ability	-0.02	-0.03	0.448	-0.07
Social Studies and Science Ability	-0.03	-0.04	0.286	-0.12
3rd Grade Year (Spring 2008)				
Language and Literacy Measures				
ECLS-K Reading	-0.18	-0.26	0.876	-0.01
PPVT (Adapted)	1.83	2.62	0.146	0.09
WJ-III Letter-Word Identification	0.44	0.63	0.818	0.01
Spanish Language and Literacy Measures				
WM Letter-Word Identification	-1.63	-2.34	0.804	-0.04
Math Skills Measures				
WJ-III Applied Problems	0.03	0.04	0.985	0.00
WJ-III Calculation	-0.05	0.07	0.960	0.00

Exhibit D.1B. Estimated IOT Impacts on Cognitive Outcomes, by Year: 3-Year-Old Cohort (continued)

Outcome	ITT Regression Adjusted Impact	IOT		
	Impact	**Impact**	**p-value**	**Effect Size**
School Performance Assessment Measures				
Promotion	-0.02	-0.03*	0.092	-0.16
Language and Literacy Ability	-0.04	-0.06	0.372	-0.13
Math Ability	0.01	0.01	0.680	0.04
Reading/Language Arts Skills++	-0.05	-0.07	0.515	-0.07
Math Skills++	0.01	0.01	0.861	0.01

Note: The three-year old cohort compliance rate is 0.69739.

Key:

******* $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

+ Indicates the reliability (Cronbach alpha) for the measure is <0.6.

++These scores are reported in the Item Response Theory (IRT) metric. Since there is not a normalizing sample to link the scores to, the scores are reported in a standardized format with a mean of 0 and a standard deviation of 1.

Bold IOT impact indicates the outcome passes the Benjamini-Hochberg test for multiple comparisons with a 10-percent false discovery rate.

Exhibit D.2A. Estimated IOT Impacts on Social-Emotional Outcomes, by Year: 4-Year-Old Cohort

Outcome	ITT Regression Adjusted Impact	IOT		
	Impact	Impact	p-value	Effect Size
Head Start Year (Spring 2003)				
Parent-Reported Measures				
Aggressive Behavior+	-0.16	-0.24	0.164	-0.15
Hyperactive Behavior+	-0.09	-0.13	0.324	-0.09
Withdrawn Behavior+	-0.04	-0.06	0.575	-0.07
Total Problem Behavior	-0.27	-0.40	0.289	-0.12
Social Competencies+	-0.04	-0.06	0.566	-0.05
Social Skills and Positive Approaches to Learning	-0.06	-0.09	0.682	-0.05
Closeness	0.25	0.37	0.146	0.14
Conflict	-0.23	-0.34	0.698	-0.05
Positive Relationships+	0.56	0.84	0.419	0.11
Kindergarten Year (Spring 2004)				
Parent-Reported Measures				
Aggressive Behavior+	-0.08	-0.12	0.477	-0.08
Hyperactive Behavior	0.11	0.16	0.273	0.11
Withdrawn Behavior+	0.00	0.00	0.986	0.00
Total Problem Behavior	0.09	0.13	0.710	0.04
Social Competencies+	-0.03	-0.05	0.770	-0.04
Social Skills and Positive Approaches to Learning+	0.07	0.10	0.483	0.07
Closeness	-0.06	-0.09	0.793	-0.03
Conflict	-0.13	-0.19	0.788	-0.03
Positive Relationships+	0.03	0.04	0.956	0.01
Teacher-Reported Measures				
ASPI – Aggressive	-0.09	-0.13	0.893	-0.02
ASPI – Inattentive/Hyperactive	-0.69	-1.03	0.286	-0.12
ASPI – Low Energy	0.10	0.15	0.888	0.02
ASPI – Oppositional	0.13	0.19	0.819	0.03
ASPI – Peer Interactions	-0.89	-1.33	0.410	-0.12
ASPI – Shy/Socially Reticent	0.64	0.96	0.418	0.13
ASPI – Structured Learning	-0.67	-1.00	0.410	-0.10
ASPI – Teacher Interaction	0.20	0.30	0.811	0.03
Closeness	0.26	0.39	0.557	0.09
Conflict	-0.35	-0.52	0.558	-0.09
Positive Relationships	0.63	0.94	0.445	0.11

Exhibit D.2A. Estimated IOT Impacts on Social-Emotional Outcomes, by Year: 4-Year-Old Cohort (continued)

Outcome	ITT Regression Adjusted Impact	IOT		
	Impact	Impact	p-value	Effect Size
1st Grade Year (Spring 2005)				
Parent-Reported Measures				
Aggressive Behavior	-0.09	-0.13	0.483	-0.08
Hyperactive Behavior	0.00	0.00	0.972	0.00
Withdrawn Behavior+	-0.13	-0.19*	0.077	-0.19
Total Problem Behavior	-0.19	-0.28	0.453	-0.07
Social Competencies+	-0.02	-0.03	0.753	-0.03
Social Skills and Positive Approaches to Learning	0.02	0.03	0.764	0.02
Closeness	-0.01	-0.01	0.944	-0.01
Conflict	-0.50	-0.75	0.373	-0.11
Positive Relationships+	0.41	0.61	0.507	0.08
Teacher-Reported Measures				
ASPI – Aggressive	-0.72	-1.08	0.257	-0.14
ASPI – Inattentive/Hyperactive	-0.26	-0.39	0.731	-0.05
ASPI – Low Energy	0.75	1.12	0.169	0.16
ASPI – Oppositional	-0.36	-0.54	0.637	-0.07
ASPI – Peer Interactions	-0.38	-0.57	0.630	-0.05
ASPI – Shy/Socially Reticent	1.37	2.05**	0.019	0.28
ASPI – Structured Learning	0.74	1.11	0.306	0.10
ASPI – Teacher Interaction	1.29	1.93*	0.099	0.19
Closeness	0.22	0.33	0.465	0.07
Conflict	0.09	0.13	0.838	0.02
Positive Relationships	0.20	0.03	0.728	0.00
3rd Grade Year (Spring 2007)				
Parent-Reported Measures				
Aggressive Behavior	-0.23	-0.34**	0.043	-0.19
Hyperactive Behavior	-0.08	-0.12	0.435	-0.07
Withdrawn Behavior+	-0.11	-0.16	0.187	-0.13
Total Problem Behavior	-0.50	-0.75*	0.090	-0.18
Social Skills and Positive Approaches to Learning	-0.10	-0.15	0.383	-0.07

Exhibit D.2A. Estimated IOT Impacts on Social-Emotional Outcomes, by Year: 4-Year-Old Cohort (continued)

Outcome	ITT Regression Adjusted Impact	IOT		
	Impact	Impact	p-value	Effect Size
Teacher-Reported Measures				
Conduct Problems-Percent in Normal Category	-0.02	-0.03	0.394	-0.09
Emotional Symptoms-Percent in Normal Category	-0.06	**-0.09*** **	0.005	-0.36
Hyperactivity-Percent in Normal Category	0.00	0.00	0.938	-0.01
Peer Problems-Percent in Normal Category	-0.04	-0.06	0.104	-0.16
Pro-social-Percent in Normal Category	-0.05	-0.07	0.163	-0.19
Total Difficulties-Percent in Normal Category	-0.06	-0.09	0.140	-0.22
Closeness with Teacher	-0.67	-1.00*	0.060	-0.19
Conflict with Teacher	0.65	0.97	0.136	0.15
Positive Teacher-Child Relationships	-1.33	-1.99*	0.063	-0.21
Social Competency++	-0.09	-0.13	0.261	-0.13
Child-Reported Measures				
Externalizing++	-0.09	-0.13	0.226	-0.13
Internalizing++	-0.10	-0.15	0.212	-0.15
Peer Relations++	-0.14	-0.21**	0.020	-0.19
School++	0.05	0.07	0.575	0.07

Note: The four-year old cohort compliance rate is 0.66897.

Key:

******* $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

+ Indicates the reliability (Cronbach alpha) for the measure is <0.6.

++These scores are reported in the Item Response Theory (IRT) metric. Since there is not a normalizing sample to link the scores to, the scores are reported in a standardized format with a mean of 0 and a standard deviation of 1.

Bold IOT impact indicates the outcome passes the Benjamini-Hochberg test for multiple comparisons with a 10-percent false discovery rate.

Exhibit D.2B. Estimated IOT Impacts on Social-Emotional Outcomes, by Year: 3-Year-Old Cohort

Outcome	ITT Regression Adjusted Impact	IOT		
	Impact	Impact	p-value	Effect Size
Head Start Year (Spring 2003)				
Parent-Reported Measures				
Aggressive Behavior	-0.10	-0.14	0.274	-0.08
Hyperactive Behavior	-0.33	**-0.47*** **	0.001	-0.30
Withdrawn Behavior+	-0.04	-0.06	0.510	-0.06
Total Problem Behavior	-0.52	**-0.75*** **	0.003	-0.20
Social Competencies+	-0.03	0.16	0.637	0.12
Social Skills and Positive Approaches to Learning	0.04	0.36	0.745	0.20
Closeness	0.18	0.19	0.220	0.07
Conflict	-0.05	-0.09	0.893	-0.01
Positive Relationships+	0.28	0.40	0.517	0.05
Age 4 Year (Spring 2004)				
Parent-Reported Measures				
Aggressive Behavior	-0.12	-0.17	0.203	-0.10
Hyperactive Behavior+	-0.13	-0.19	0.242	-0.13
Withdrawn Behavior+	-0.08	-0.11	0.248	-0.12
Total Problem Behavior	-0.39	-0.56	0.115	-0.15
Social Competencies	-0.01	0.16	0.868	0.12
Social Skills and Positive Approaches to Learning	0.19	0.27*	0.055	0.16
Closeness	0.22	0.32	0.232	0.12
Conflict	-0.39	-0.56	0.461	-0.08
Positive Relationships+	0.72	1.03	0.261	0.13
Kindergarten Year (Spring 2005)				
Parent-Reported Measures				
Aggressive Behavior	-0.08	-0.11	0.382	-0.06
Hyperactive Behavior	-0.18	-0.26**	0.048	-0.17
Withdrawn Behavior+	-0.03	-0.04	0.563	-0.04
Total Problem Behavior	-0.26	-0.37	0.246	-0.10
Social Competencies+	0.11	0.16	0.179	0.12
Social Skills and Positive Approaches to Learning	0.25	0.36*	0.075	0.20
Closeness	0.13	0.19	0.434	0.07
Conflict	-0.06	-0.09	0.888	-0.01
Positive Relationships	0.12	0.17	0.803	0.02

Exhibit D.2B. Estimated IOT Impacts on Social-Emotional Outcomes, by Year: 3-Year-Old Cohort (continued)

Outcome	ITT Regression Adjusted Impact	IOT		
	Impact	Impact	p-value	Effect Size
Teacher-Reported Measures				
ASPI – Aggressive	0.40	0.57	0.318	0.08
ASPI – Inattentive/Hyperactive	-0.02	-0.03	0.972	0.00
ASPI – Low Energy	0.52	0.75	0.241	0.11
ASPI – Oppositional	0.03	0.04	0.953	0.01
ASPI – Peer Interactions	0.64	0.92	0.335	0.08
ASPI – Shy/Socially Reticent	0.07	0.10	0.882	0.01
ASPI – Structured Learning	0.74	10.6	0.136	0.10
ASPI – Teacher Interaction	0.27	0.39	0.649	0.04
Closeness	-0.26	-0.37	0.295	-0.08
Conflict	0.03	0.04	0.959	0.01
Positive Relationships	-0.40	-0.57	0.483	-0.06
1st Grade Year (Spring 2006)				
Parent-Reported Measures				
Aggressive Behavior	-0.05	-0.07	0.624	-0.04
Hyperactive Behavior	-0.11	-0.16	0.127	-0.10
Withdrawn Behavior+	0.02	0.03	0.732	0.03
Total Problem Behavior	-0.15	-0.22	0.439	-0.05
Social Competencies	0.08	0.11	0.317	0.09
Social Skills and Positive Approaches to Learning	0.05	0.07	0.642	0.04
Closeness	0.29	0.42**	0.013	0.15
Conflict	-0.55	-0.79	0.210	-0.12
Positive Relationships	0.77	1.10*	0.098	0.14
Teacher-Reported Measures				
ASPI – Aggressive	-0.54	-0.77	0.266	-0.10
ASPI – Inattentive/Hyperactive	-0.45	-0.65	0.402	-0.08
ASPI – Low Energy	0.44	0.63	0.383	0.09
ASPI – Oppositional	0.04	0.06	0.944	0.01
ASPI – Peer Interactions	-0.43	-0.62	0.584	-0.05
ASPI – Shy/Socially Reticent	0.21	0.30	0.623	0.04
ASPI – Structured Learning	-0.12	-0.17	0.855	-0.02
ASPI – Teacher Interaction	-0.15	-0.22	0.819	-0.02
Closeness	0.28	0.40	0.301	0.09
Conflict	-0.16	-0.23	0.722	-0.03
Positive Relationships	0.44	0.63	0.461	0.07

Exhibit D.2B. Estimated IOT Impacts on Social-Emotional Outcomes, by Year: 3-Year-Old Cohort (continued)

Outcome	ITT Regression Adjusted Impact	IOT		
	Impact	Impact	p-value	Effect Size
3rd Grade Year (Spring 2008)				
Parent-Reported Measures				
Aggressive Behavior	0.04	0.06	0.703	0.03
Hyperactive Behavior	0.01	0.01	0.942	0.00
Withdrawn Behavior	0.10	0.14	0.230	0.11
Total Problem Behavior	0.15	0.21	0.604	0.04
Social Skills and Positive Approaches to Learning	0.24	0.34**	0.025	0.17
Teacher-Reported Measures				
Conduct Problems-Percent in Normal Category	-0.01	-0.01	0.759	-0.03
Emotional Symptoms-Percent in Normal Category	0.01	0.01	0.705	0.04
Hyperactivity-Percent in Normal Category	0.02	0.03	0.431	0.07
Peer Problems-Percent in Normal Category	0.00	0.00	0.984	0.00
Pro-social-Percent in Normal Category	-0.05	-0.07	0.137	-0.17
Total Difficulties-Percent in Normal Category	0.01	0.01	0.832	0.01
Closeness with Teacher	-0.40	-0.57	0.300	-0.11
Conflict with Teacher	0.27	0.39	0.613	0.06
Positive Teacher-Child Relationships	-0.45	-0.64	0.549	-0.06
Social Competency++	-0.05	-0.07	0.427	-0.07
Child-Reported Measures				
Externalizing++	-0.02	-0.03	0.733	-0.03
Internalizing++	0.02	0.03	0.731	0.03
Peer Relations++	0.08	0.11	0.227	0.13
School++	0.05	0.07	0.564	0.06

Note: The three-year old cohort compliance rate is 0.69739.

Key:

******* $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

+ Indicates the reliability (Cronbach alpha) for the measure is <0.6.

++ These scores are reported in the Item Response Theory (IRT) metric. Since there is not a normalizing sample to link the scores to, the scores are reported in a standardized format with a mean of 0 and a standard deviation of 1.

Bold IOT impact indicates the outcome passes the Benjamini-Hochberg test for multiple comparisons with a 10-percent false discovery rate.

Exhibit D.3A. Estimated IOT Impacts on Parent-Reported Health Outcomes, by Year: 4-Year-Old Cohort

Outcome	ITT Regression Adjusted Impact	IOT		
	Impact	Impact	p-value	Effect Size
Head Start Year (Spring 2003)				
Parent-Reported Measures				
Child Received Dental Care	0.15	**0.22*** **	0.000	0.45
Child Has Health Insurance Coverage	0.01	0.01	0.733	0.05
Child's Overall Health Status is Excellent/Good	-0.03	-0.04	0.244	-0.12
Child Needs Ongoing Care	0.01	0.01	0.422	0.05
Child Had Care for Injury Last Month	-0.02	-0.03	0.409	-0.09
Kindergarten Year (Spring 2004)				
Parent-Reported Measures				
Child Received Dental Care	0.03	0.04	0.435	0.09
Child Has Health Insurance Coverage	0.04	0.06*	0.056	0.17
Child's Overall Health Status is Excellent/GOod	0.05	0.07*	0.098	0.18
Child Needs Ongoing Care	-0.02	-0.03	0.432	-0.09
Child Had Care for Injury Last Month	0.02	0.03	0.547	0.10
1st Grade Year (Spring 2005)				
Parent-Reported Measures				
Child Received Dental Care	0.02	0.03	0.550	0.06
Child Has Health Insurance Coverage	0.04	0.06**	0.044	0.17
Child's Overall Health Status is Excellent/Good	-0.01	-0.01	0.858	-0.04
Child Needs Ongoing Care	0.02	0.03	0.291	0.09
Child Had Care for Injury Last Month	0.02	0.03	0.303	0.09

Exhibit D.3A. Estimated IOT Impacts on Parent-Reported Health Outcomes, by Year: 4-Year-Old Cohort (continued)

Outcome	ITT Regression Adjusted Impact	IOT		
	Impact	Impact	p-value	Effect Size
3rd Grade Year (Spring 2007)				
Parent-Reported Measures				
Child Received Dental Care	0.01	0.01	0.730	0.04
Child Has Health Insurance Coverage	0.00	0.00	1.000	0.00
Child's Overall Health Status is Excellent/Good	0.01	0.01	0.547	0.04
Child Needs Ongoing Care	0.00	0.00	0.861	0.01
Child Had Care for Injury Last Month	0.01	0.01	0.815	0.03

Note: The four-year old cohort compliance rate is 0.66897.

Key:

******* $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

Bold IOT impact indicates the outcome passes the Benjamini-Hochberg test for multiple comparisons with a 10-percent false discovery rate.

Exhibit D.3B. Estimated IOT Impacts on Parent-Reported Health Outcomes, by Year: 3-Year-Old Cohort

Outcome	ITT Regression Adjusted Impact	IOT		
	Impact	Impact	p-value	Effect Size
Head Start Year (Spring 2003)				
Parent-Reported Measures				
Child Received Dental Care	0.17	**0.24*** **	0.000	0.49
Child Has Health Insurance Coverage	0.00	0.00	0.803	0.00
Child's Overall Health Status is Excellent/Good	0.05	0.07**	0.045	0.17
Child Needs Ongoing Care	0.00	0.00	0.988	0.00
Child Had Care for Injury Last Month	-0.01	-0.01	0.699	-0.05
Age 4 Year (Spring 2004)				
Parent-Reported Measures				
Child Received Dental Care	0.10	**0.14*** **	0.001	0.30
Child Has Health Insurance Coverage	0.00	0.00	0.935	0.00
Child's Overall Health Status is Excellent/Good	0.00	0.00	0.851	0.00
Child Needs Ongoing Care	0.01	0.01	0.739	0.04
Child Had Care for Injury Last Month	0.03	0.04	0.089	0.15
Kindergarten Year (Spring 2005)				
Parent-Reported Measures				
Child Received Dental Care	0.03	0.04	0.270	0.10
Child Has Health Insurance Coverage	0.04	0.06**	0.044	0.19
Child's Overall Health Status is Excellent/Good	0.00	0.00	0.889	0.00
Child Needs Ongoing Care	-0.03	-0.04	0.114	-0.11
Child Had Care for Injury Last Month	0.00	0.00	0.985	0.00
1st Grade Year (Spring 2006)				
Parent-Reported Measures				
Child Received Dental Care	0.01	0.01	0.786	0.03
Child Has Health Insurance Coverage	0.02	0.03	0.252	0.10
Child's Overall Health Status is Excellent/Good	0.02	0.03	0.434	0.08
Child Needs Ongoing Care	-0.01	-0.01	0.578	-0.04
Child Had Care for Injury Last Month	0.02	0.03	0.294	0.11

Exhibit D.3B. Estimated IOT Impacts on Parent-Reported Health Outcomes, by Year: 3-Year-Old Cohort (continued)

Outcome	ITT Regression Adjusted Impact	IOT		
	Impact	Impact	p-value	Effect Size
3rd Grade Year (Spring 2008)				
Parent-Reported Measures				
Child Received Dental Care	0.00	0.00	0.979	0.00
Child Has Health Insurance Coverage	0.02	0.03	0.354	0.09
Child's Overall Health Status is Excellent/Good	-0.01	-0.01	0.681	-0.03
Child Needs Ongoing Care	0.02	0.03	0.421	0.09
Child Had Care for Injury Last Month	-0.04	-0.06	0.216	-0.16

Note: The three-year old cohort compliance rate is 0.69739.

Key:

******* $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

Bold IOT impact indicates the outcome passes the Benjamini-Hochberg test for multiple comparisons with a 10-percent false discovery rate.

Exhibit D.4A. Estimated IOT Impacts on Parenting Outcomes, by Year: 4-Year-Old Cohort

Outcome	ITT Regression Adjusted Impact	IOT		
	Impact	Impact	p-value	Effect Size
Head Start Year (Spring 2003)				
Parent-Reported Measures				
Parent Spanked Child in Last Week	-0.01	-0.01	0.750	-0.03
Parent Used Time Out in Last Week	-0.08	-0.12**	0.025	-0.26
Parent Read to Child in Last Week	0.03	0.04	0.396	0.10
Parental Safety Practices Scale+	0.03	0.04	0.382	0.00
Family Cultural Enrichment Scale+	0.08	0.12	0.368	0.08
Kindergarten Year (Spring 2004)				
Parent-Reported Measures				
Parent Spanked Child in Last Week	0.00	0.00	0.869	0.00
Parent Used Time Out in Last Week	0.01	0.01	0.689	0.03
Parent Read to Child in Last Week	-0.03	-0.04	0.385	-0.09
Parental Safety Practices Scale+	0.04	0.06	0.156	0.16
Family Cultural Enrichment Scale+	0.14	0.21	0.142	0.15
Parent Style: Authoritarian	0.00	0.00	0.981	0.00
Parent Style: Authoritative	0.05	0.07	0.164	0.15
Parent Style: Neglectful	-0.03	-0.04	0.211	-0.14
Parent Style: Permissive	-0.02	-0.03	0.447	-0.08
Teacher-Reported Measures				
School Contact and Communication	-0.01	-0.01	0.845	-0.04
Parent Participation	-0.01	-0.01	0.841	-0.05
1st Grade Year (Spring 2005)				
Parent-Reported Measures				
Parent Spanked Child in Last Week	0.00	0.00	0.976	0.00
Parent Used Time Out in Last Week	-0.04	-0.06	0.322	-0.12
Parent Read to Child in Last Week	-0.01	-0.01	0.733	-0.03
Family Cultural Enrichment Scale+	0.04	0.06	0.612	0.04
Parent Style: Authoritarian	-0.03	-0.04	0.199	-0.15
Parent Style: Authoritative	0.04	0.06	0.158	0.12
Parent Style: Neglectful	-0.02	-0.03	0.327	-0.11
Parent Style: Permissive	0.00	0.00	0.936	0.00
Teacher-Reported Measures				
School Contact and Communication	-0.02	-0.03	0.570	-0.08
Parent Participation	-0.01	-0.01	0.817	-0.04

Exhibit D.4A. Estimated IOT Impacts on Parenting Outcomes, by Year: 4-Year-Old Cohort (continued)

Outcome	ITT Regression Adjusted Impact	IOT		
	Impact	Impact	p-value	Effect Size
3rd Grade Year (Spring 2007)				
Parent-Reported Measures				
Parent Spanked Child in Last Week	-0.02	-0.03	0.635	-0.06
Parent Style: Authoritarian	0.02	0.03	0.494	0.07
Parent Style: Authoritative	-0.02	-0.03	0.516	-0.07
Parent Style: Neglectful	0.00	0.00	0.843	-0.01
Parent Style: Permissive	0.01	0.01	0.523	0.06
Supportive School Environment++	0.03	0.04	0.701	0.04
Effect of Parenting on Parent's Life++	0.06	0.09	0.450	0.09
Doing Things Together++	-0.02	-0.03	0.786	-0.03
Time Spent with Child++	0.27	**0.40*** **	0.001	0.40
Parent Perception of School Services++	-0.09	-0.13	0.175	-0.15
Teacher-Reported Measures				
School Contact and Communication	-0.04	-0.06	0.103	-0.15
Parent Participation	-0.01	-0.01	0.652	-0.06

Note: The four-year old cohort compliance rate is 0.66897.

Key:

*** $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

+ Indicates the reliability (Cronbach alpha) for the measure is <0.6.

++These scores are reported in the Item Response Theory (IRT) metric. Since there is not a normalizing sample to link the scores to, the scores are reported in a standardized format with a mean of 0 and a standard deviation of 1.

Bold IOT impact indicates the outcome passes the Benjamini-Hochberg test for multiple comparisons with a 10-percent false discovery rate.

Exhibit D.4B. Estimated IOT Impacts on Parenting Outcomes, by Year: 3-Year-Old Cohort

Outcome	ITT Regression Adjusted Impact	IOT		
	Impact	Impact	p-value	Effect Size
Head Start Year (Spring 2003)				
Parent-Reported Measurers				
Parent Spanked Child in Last Week	-0.07	**-0.10****	0.025	-0.20
Parent Used Time Out in Last Week	-0.04	-0.06	0.205	-0.12
Parent Read to Child in Last Week	0.07	**0.10****	0.030	0.22
Parental Safety Practices Scale+	0.03	0.04	0.146	0.13
Family Cultural Enrichment Scale+	0.25	**0.36*****	0.000	0.26
Age 4 Year (Spring 2004)				
Parent-Reported Measurers				
Parent Spanked Child in Last Week	0.01	0.01	0.635	0.03
Parent Used Time Out in Last Week	-0.02	-0.03	0.355	-0.06
Parent Read to Child in Last Week	-0.01	-0.01	0.827	-0.03
Parental Safety Practices Scale+	0.02	0.03	0.313	0.08
Family Cultural Enrichment Scale+	0.04	0.06	0.593	0.04
Parent Style: Authoritarian	-0.04	**-0.06*****	0.005	-0.21
Parent Style: Authoritative	0.04	0.06	0.186	0.12
Parent Style: Neglectful	0.00	0.00	0.826	0.00
Parent Style: Permissive	0.00	0.00	0.893	0.00
Kindergarten Year (Spring 2005)				
Parent-Reported Measures				
Parent Spanked Child in Last Week	-0.04	-0.06*	0.070	-0.12
Parent Used Time Out in Last Week	-0.07	-0.10**	0.013	-0.20
Parent Read to Child in Last Week	0.03	0.04	0.305	0.09
Parental Safety Practices Scale+	0.01	0.01	0.714	0.04
Family Cultural Enrichment Scale+	0.00	0.00	0.968	0.00
Parent Style: Authoritarian	0.00	0.00	0.950	0.00
Parent Style: Authoritative	0.00	0.00	0.905	0.00
Parent Style: Neglectful	-0.02	-0.03	0.202	-0.11
Parent Style: Permissive	0.03	0.04	0.310	0.11
Teacher-Reported Measures				
School Contact and Communication	0.00	0.00	0.879	0.00
Parent Participation	0.00	0.00	0.886	0.00

Exhibit D.4B. Estimated IOT Impacts on Parenting Outcomes, by Year: 3-Year-Old Cohort (continued)

Outcome	ITT Regression Adjusted Impact	IOT		
	Impact	Impact	p-value	Effect Size
1st Grade Year (Spring 2006)				
Parent-Reported Measures				
Parent Spanked Child in Last Week	-0.03	-0.04	0.183	-0.10
Parent Used Time Out in Last Week	-0.05	-0.07*	0.075	-0.14
Parent Read to Child in Last Week	0.01	0.01	0.746	0.03
Family Cultural Enrichment Scale+	0.01	0.01	0.879	0.01
Parent Style: Authoritarian	-0.03	-0.04**	0.046	-0.16
Parent Style: Authoritative	0.00	0.00	0.877	0.00
Parent Style: Neglectful	0.00	0.00	0.974	0.00
Parent Style: Permissive	0.02	0.03	0.320	0.08
Teacher-Reported Measures				
School Contact and Communication	0.02	0.03	0.453	0.07
Parent Participation	0.01	0.01	0.626	0.04
3rd Grade Year (Spring 2008)				
Parent-Reported Measures				
Parent Spanked Child in Last Week	-0.02	-0.03	0.630	-0.06
Parent Style: Authoritarian	-0.03	-0.04	0.298	-0.11
Parent Style: Authoritative	0.08	0.11**	0.033	0.23
Parent Style: Neglectful	-0.01	-0.01	0.495	-0.07
Parent Style: Permissive	-0.04	-0.06	0.160	-0.14
Supportive School Environment++	0.03	0.04	0.672	0.04
Effect of Parenting on Parent's Life++	-0.04	-0.06	0.429	-0.06
Doing Things Together++	0.08	0.11	0.160	0.13
Time Spent with Child++	-0.05	-0.07	0.408	-0.07
Parent Perception of School Services++	0.11	0.16	0.142	0.16
Teacher-Reported Measures				
School Contact and Communication	0.00	0.00	0.883	-0.01
Parent Participation	-0.03	-0.04	0.181	

Note: The three-year old cohort compliance rate is 0.69739.

Key:

******* $p \leq 0.01$

** $p \leq 0.05$

* $p \leq 0.10$

+ Indicates the reliability (Cronbach alpha) for the measure is <0.6.

++These scores are reported in the Item Response Theory (IRT) metric. Since there is not a normalizing sample to link the scores to, the scores are reported in a standardized format with a mean of 0 and a standard deviation of 1.

Bold IOT impact indicates the outcome passes the Benjamini-Hochberg test for multiple comparisons with a 10-percent false discovery rate.

Appendix E

Subgroup Tables, 2003-2008

Exhibit E.1. Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Child's Pre-Academic Skills

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) Children not in Lowest Quartile	Impact in Subgroup B (Head Start –Control) Children in Lowest Quartile	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE				
Direct Child Assessment				
Head Start Year (11)	--	--	--	--
Kindergarten (12)	WJ III Oral Comprehension	0.19	-4.24**	A-B**
	WJ III Math Reasoning	0.96	-3.28*	A-B*
1st grade (15)	WJ III Applied Problems	2.36*	-3.85	A-B*
3rd grade (5)	--	--	--	--
School Performance				
Head Start Year (0)				
Kindergarten (5)	--	--	--	--
1st grade (5)	Math Ability	-0.01	-0.15**	A-B*
3rd grade (5)	--	--	--	---
SOCIAL-EMOTIONAL				
Parent Report				
Head Start Year (9)	Closeness	-0.04	1.08**	A-B**
	Positive Relationships	-0.16	2.60*	A-B*
Kindergarten (9)	--	--	--	--
1st grade (9)	--	--	--	--
3rd grade (5)	--	--	--	--
Teacher Report				
Head Start Year (0)				
Kindergarten (11)	ASPI-Oppositional	0.74	-1.65*	A-B*
	ASPI-Problems with Peer Interaction	0.22	-4.17***	A-B**
	Conflict	0.38	-2.52*	A-B**
	Positive Relationships	-0.29	3.38**	A-B**
1st grade (11)	Closeness	-0.17	1.36*	A-B*
	Positive Relationships	-0.57	2.44*	A-B*
3rd grade (10)	--	--	--	--
Child Report				
3rd grade (4)	Peer Relations	**-0.21*** **	0.08	A-B*

Exhibit E.1. Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Child's Pre-Academic Skills (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) Children not in Lowest Quartile	Impact in Subgroup B (Head Start –Control) Children in Lowest Quartile	Statistically Significant Differences in Impacts Between Subgroups
HEALTH				
Head Start Year (5)	--	--	--	--
Kindergarten (5)	--	--	--	--
1st grade (5)	--	--	--	--
3rd grade (5)	--	--	--	--
PARENTING				
Parent Report				
Head Start Year (5)	--	--	--	--
Kindergarten (9)	Parent Spanked Child in Last Week	0.04	**-0.13*** **	A-B**
	Parent Used Time Out in Last Week	-0.02	0.10*	A-B*
1st grade (8)	Parent Read to Child in Last Week	0.03	-0.12*	A-B*
3rd grade (10)	Supportive School Environment	0.14*	**-0.30** **	**A-B *** **
	Parenting Style: Neglectful	0.03	**-0.12** **	**A-B *** **
	Parent Perception of School Services	-0.01	**-0.31** **	A-B*
Teacher Report				
Head Start Year (0)				
Kindergarten (2)	--	--	--	--
1st grade (2)	--	--	--	--
3rd grade (2)	--	--	--	--

Key:
*** $p \leq 0.01$
** $p \leq 0.05$
* $p \leq 0.10$
Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.
Gray cell indicates no data available.

Exhibit E.2. Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Child's Home Language

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) Dual Language Learners	Impact in Subgroup B (Head Start –Control) Children English Speaking	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE				
Direct Child Assessment				
Head Start Year (11)	PPVT (Adapted)	**8.97*** **	1.23	**A-B** **
Kindergarten (12)	--	--		--
1st grade (15)	--	--	--	--
3rd grade (5)	--	--	--	--
School Performance				
Head Start Year (0)				
Kindergarten (5)	Math Ability	0.15**	0.00	A-B*
1st grade (5)	--	--	--	--
3rd grade (5)	--	--	--	--
SOCIAL-EMOTIONAL				
Parent Report				
Head Start Year (9)	--	--	--	--
Kindergarten (9)	--	--	--	--
1st grade (9)	--	--	--	--
3rd grade (5)	--	--	--	--
Teacher Report				
Head Start Year (0)				
Kindergarten (11)	--	--	--	--
1st grade (11)	--	--	--	--
3rd grade (10)	--	--	--	--
Child Report				
3rd grade (4)	--	--	--	--
HEALTH				
Head Start Year (5)	Child's Overall Health Status is Excellent/ Good Child Has Health Insurance Coverage	**-0.08*** ** 0.07*	-0.01 -0.02	A-B* A-B*
Kindergarten (5)	Child Has Health Insurance Coverage	0.11**	0.01	A-B*
1st grade (5)	Child Received Dental Care	0.08*	-0.01	A-B*
3rd grade (5)	--	--	--	--

Exhibit E.2. Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Child's Home Language (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) Dual Language Learners	Impact in Subgroup B (Head Start –Control) Children English Speaking	Statistically Significant Differences in Impacts Between Subgroups
PARENTING				
Parent Report				
Head Start Year (5)	Parental Safety Practices Scale	0.09*	-0.00	A-B*
Kindergarten (9)	--	--	--	--
1st grade (8)	Parenting Style: Neglectful	-0.06**	0.00	A-B*
3rd grade (10)	Supportive School Environment	**0.35*** **	-0.11	**A-B*** **
Teacher Report				
Head Start Year (0)				
Kindergarten (2)	--	--	--	--
1st grade (2)	--	--	--	--
3rd grade (2)	--	--	--	--

Key:
*** $p \leq 0.01$
** $p \leq 0.05$
* $p \leq 0.10$
Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.
Gray cell indicates no data available.

Exhibit E.3. Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Special Needs

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) Children with No Special Needs	Impact in Subgroup B (Head Start –Control) Children with Special Needs	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE				
Direct Child Assessment				
Head Start Year (11)	--	--	--	--
Kindergarten (12)	--	--	--	--
1st grade (15)	--	--	--	--
3rd grade (5)	WJIII Letter-Word Identification	3.71*	-8.66	A-B*
School Performance				
Head Start Year (0)				
Kindergarten (5)	--	--	--	--
1st grade (5)	--	--	--	--
3rd grade (5)	--	--	--	--
SOCIAL-EMOTIONAL				
Parent Report				
Head Start Year (9)	--	--	--	--
Kindergarten (9)	Social Competencies	0.04	-0.46**	A-B***
1st grade (9)	--	--	--	--
3rd grade (5)				
Teacher Report				
Head Start Year (0)				
Kindergarten (11)	ASPI – Problems with Structured Learning	-1.45*	4.60*	A-B**
1st grade (11)	--	--	--	--
Child Report				
3rd grade (4)	--	--	--	--
HEALTH				
Head Start Year (5)	--	--	--	--
Kindergarten (5)	Child's Overall Health Status Is Excellent/ Good	0.04	**0.17****	A-B*
	Child Had Care for Injury Last Month	0.00	**0.09****	A-B*
1st grade (5)	Child Needs Ongoing Care	0.00	0.16*	A-B*
	Child Had Care for Injury Last Month	-0.01	**0.22*****	**A-B*****
3rd grade (5)	--	--	--	--

Exhibit E.3. Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Special Needs (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) Children with No Special Needs	Impact in Subgroup B (Head Start –Control) Children with Special Needs	Statistically Significant Differences in Impacts Between Subgroups
PARENTING				
Parent Report				
Head Start Year (5)	--	--	--	--
Kindergarten (9)	Parenting Style: Neglectful	-0.01	-0.14*	A-B*
1st grade (8)	--	--	--	--
3rd grade (10)	--	--	--	--
Teacher Report				
Head Start Year (0)				
Kindergarten (2)	--	--	--	--
1st grade (2)	--	--	--	--
3rd grade (2)	--	--	--	--

Key:
******* $p \leq 0.01$
** $p \leq 0.05$
***** $p \leq 0.10$
Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.
Gray cell indicates no data available.

Exhibit E.4. Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Biological Mother/ Caregiver Race/Ethnicity

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) White/Other Children	Impact in Subgroup B (Head Start – Control) Black Children	Impact in Subgroup C (Head Start – Control) Hispanic Children	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE					
Direct Child Assessment					
Head Start Year (11)	PPVT (Adapted)	0.90	-2.04	**9.35*** **	A-C**
					B-C**
	Counting Bears	-0.06	--	0.09*	A-C*
	WJ III Spelling	1.85	**8.62*** **	--	A-B*
Kindergarten (12)	WJ III Spelling	-4.90*	6.62**	-1.55	**A-B*** **
					B-C*
	WJ III Basic Reading Skills	-2.84	5.95*	-3.44	A-B*
					B-C**
1st grade (15)	--	--	--	--	--
3rd grade (5)	--	--	--	--	--
School Performance					
Head Start Year (0)					
Kindergarten (5)	--	--	--	--	--
1st grade (5)	--	--	--	--	--
3rd grade (5)	--	--	--	--	--
SOCIAL-EMOTIONAL					
Parent Report					
Head Start Year (9)	--	--	--	--	--
Kindergarten (9)	--	--	--	--	--
1st grade (9)	Social Skills and Positive Approaches to Learning	0.22	0.33	-0.32***	A-C**
					B-C***
3rd grade (5)	Aggressive Behavior	0.09	**-0.76*** **	-0.16	A-B**
					B-C**
	Hyperactive Behavior		-0.47**	0.14	B-C**
	Total Behavior Problems	-0.13	**-1.59*** **	-0.11	A-B*
					B-C**

Exhibit E.4. Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Biological Mother/ Caregiver Race/Ethnicity (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) White/Other Children	Impact in Subgroup B (Head Start – Control) Black Children	Impact in Subgroup C (Head Start – Control) Hispanic Children	Statistically Significant Differences in Impacts Between Subgroups
Teacher Report					
Head Start Year (0)					
Kindergarten (11)	ASPI-Inattentive/Hyperactive	-0.36	-4.60**	0.86	A-B** B-C***
	ASPI-Problems with Structured Learning	1.02	-5.34**	0.07	A-B*** B-C**
	ASPI-Problems with Peer Interaction	1.47	-6.12*	--	A-B**
	Positive Relationships	-0.56	4.07*	0.08	A-B* B-C*
	ASPI-Problems with Teacher Interaction	1.85	-3.37*	0.38	A-B** B-C*
	ASPI-Shy/Socially Reticent	2.25*	--	-0.84	A-C*
1st grade (11)	ASPI-Withdrawn/Low Energy	2.25**	-1.51	--	A-B**
	ASPI-Shy/Socially Reticent	2.73***	--	0.67	A-C*
	ASPI-Problems with Teacher Interaction	3.12**	--	0.08	A-C**
3rd grade (10)	Closeness with Teacher	**-1.60*** **	--	0.18	A-C**
Child Report					
3rd grade (4)	Internalizing	0.04	--	-0.26**	A-C**
	Peer Relations	**-0.32*** **	--	-0.02	A-C**
	School	-0.08	0.31*	--	A-B**
HEALTH					
Head Start Year (5)	Child's Overall Health Status Is Excellent/Good	**-0.09****	0.03	--	A-B**
Kindergarten (5)	--	--	--	--	--
1st grade (5)	Child's Overall Health Status Is Excellent/Good	-0.03	0.10*	-0.05	A-B* B-C*
3rd grade (5)	--	--	--	--	--

Exhibit E.4. Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Biological Mother/ Caregiver Race/Ethnicity (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) White/Other Children	Impact in Subgroup B (Head Start – Control) Black Children	Impact in Subgroup C (Head Start – Control) Hispanic Children	Statistically Significant Differences in Impacts Between Subgroups
PARENTING					
Parent Report					
Head Start Year (5)	--	--	--	--	--
Kindergarten (9)	Parenting Style: Neglectful	-0.09**	--	-0.00	A-C*
	Parent Read to Child in Last Week	-0.10*	0.07	--	A-B*
	Family Cultural Enrichment Scale	--	-0.15	0.34**	B-C*
1st grade (8)	Parenting Style: Neglectful	--	0.05	-0.05**	**B-C****
	Parent Spanked Child in Last Week	0.03	**-0.11*****	0.05	**A-B*****
					B-C***
	Parenting Style: Permissive	0.03	-0.10*	0.04	A-B*
					B-C*
3rd Grade (10)	Supportive School Environment	--	-0.20	0.20**	B-C**
	Parent Spanked Child in Last Week	0.04	-0.22**	0.06	A-B**
					B-C**
Teacher Report					
Head Start Year (0)					
Kindergarten (2)	--	--	--	--	--
1st grade (2)	--	--	--	--	--
3rd grade (2)	Parent Participation	-0.08*	--	0.04	A-C*

Key:
*** $p \leq 0.01$
** $p \leq 0.05$
* $p \leq 0.10$
Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.
Gray cell indicates no data available.

Exhibit E.5. Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Parent/Caregiver Reported Depressive Symptoms

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) No Symptoms	Impact in Subgroup B (Head Start – Control) Mild Symptoms	Impact in Subgroup C (Head Start – Control) Moderate Symptoms	Impact in Subgroup D (Head Start – Control) Severe Symptoms	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE						
Direct Child Assessment						
Head Start Year (11)	PPVT (Adapted)	5.30**	0.20	--	--	A-B*
	Counting Bears	--	**0.16****	-0.06	--	B-C*
	Letter Naming	0.69	**3.90*****	**4.75*****	**4.64*****	A-B* **A-C***** A-D**
	WJ III Pre-Academic Skills	1.06	**5.68****	--	--	A-B*
	WJ III Letter Word Identification	2.90	**7.26****	--	**13.61*****	**A-D***** B-D*
	Color Identification	0.02	--	--	0.20**	A-D*
Kindergarten (12)	--	--	--	--	--	--
1st grade (15)	WJ III Academic Application	-0.98	3.88*	--	-3.23	A-B* B-D**
	WJ III Quantitative Concepts	1.13	0.56	--	-4.81*	A-D* B-D*
	WJ III Passage Comprehension	-2.60	6.37*	-5.82	-3.44	A-B** B-C* B-D**
	WJ III Applied Problems	--	3.87**	--	-3.96	B-D**
	WJ III Word Attack	--	7.56*	--	-5.55	B-D*
3rd grade (5)	ECLS-K Reading	-0.11	5.24**	--	--	A-B*
School Performance						
Head Start Year (0)						
Kindergarten (5)	--	--	--	--	--	--
1st grade (5)	--	--	--	--	--	--
3rd grade (5)	Promotion	-0.02	--	0.11*	--	A-C*

Exhibit E.5. Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Parent/Caregiver Reported Depressive Symptoms (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) No Symptoms	Impact in Subgroup B (Head Start – Control) Mild Symptoms	Impact in Subgroup C (Head Start – Control) Moderate Symptoms	Impact in Subgroup D (Head Start – Control) Severe Symptoms	Statistically Significant Differences in Impacts Between Subgroups
SOCIAL-EMOTIONAL						
Parent Report						
Head Start Year (9)	Hyperactive Behavior	0.02	**-0.46****	--	--	A-B**
	Total Problem Behavior	-0.11	**-1.05*****	0.70	--	A-B* B-C**
Kindergarten (9)	Withdrawn Behavior	-0.18**	0.19*	0.36	-0.38	**A-B***** A-C** B-D**
	Social Competencies	0.11	--	--	-0.48**	A-D**
1st grade (9)	Social Competencies	0.03	-0.32***	--	--	A-B**
	Aggressive Behavior	--	0.02	0.20	-0.75*	B-D* C-D**
3rd grade (5)	Aggressive Behavior	-0.08	--	0.20	-1.20**	A-D** C-D**
	Hyperactive Behavior	0.09	--	0.26	-0.69*	A-D* C-D*
	Total Problem Behavior	-0.25	-0.50	0.52	-2.21**	A-D* B-D* C-D*
Teacher Report						
Head Start Year (0)						
Kindergarten (11)	ASPI-Inattentive/ Hyperactive	0.74	-2.63*	--	--	A-B*
1st grade (11)	ASPI–Withdrawn/Low Energy	-0.18	0.21	4.23**	--	A-C** B-C*
	ASPI–Aggressive	0.05	-0.16	--	-3.98**	A-D* B-D*

Exhibit E.5. Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Parent/Caregiver Reported Depressive Symptoms (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) No Symptoms	Impact in Subgroup B (Head Start – Control) Mild Symptoms	Impact in Subgroup C (Head Start – Control) Moderate Symptoms	Impact in Subgroup D (Head Start – Control) Severe Symptoms	Statistically Significant Differences in Impacts Between Subgroups
3rd grade	Prosocial Behavior+	-0.04	**-0.18*** **	0.07	0.08	A-B* B-C* B-D*
	Closeness with Teacher	--	-1.39*	--	1.43	B-D**
	Conflict with Teacher	--	2.13**	--	-3.11	B-D**
	Positive Relationship with Teacher	--	-3.40**	--	5.41	B-D**
	Total Difficulties+	--	-0.15*	-0.20*	0.20	B-D** C-D*
	Social Competency	-0.15*	--	--	0.38	A-D*
Child Report						
3rd grade (4)	Internalizing	-0.20*	--	--	0.35	A-D***
HEALTH						
Head Start Year (5)	Child Needs Ongoing Care	0.02	-0.00	--	0.10**	A-D* B-D**
Kindergarten (5)	Child Had Care for Injury Last Month	--	--	0.14**	-0.07	C-D*
	Child Received Dental Care	0.02	-0.06	--	0.21**	A-D* B-D*
1st grade (5)	Child Needs Ongoing Care	0.01	--	--	**0.13****	A-D*
	Child Had Care for Injury Last Month	--	-0.05	-0.00	**0.16****	B-D** C-D*
3rd grade (5)	--	--	--	--	--	--
PARENTING						
Parent Report						
Head Start Year (5)	--	--	--	--	--	--
Kindergarten (9)	Parenting Style: Neglectful	-0.04*	--	0.08	-0.17*	A-C** C-D**

Exhibit E.5. Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Parent/Caregiver Reported Depressive Symptoms (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) No Symptoms	Impact in Subgroup B (Head Start – Control) Mild Symptoms	Impact in Subgroup C (Head Start – Control) Moderate Symptoms	Impact in Subgroup D (Head Start – Control) Severe Symptoms	Statistically Significant Differences in Impacts Between Subgroups
1st grade (8)	Parenting Style: Authoritarian	--	-0.01	0.02	-0.15*	B-D* C-D**
3rd grade (10)	Parenting Style: Neglectful	0.03	0.05	**-0.24*** **	--	**A-C***** **B-C*****
	Parenting Style: Permissive	-0.05	--	0.15*	--	A-C**
	Effect of Parenting on Parent's Life	0.14*	--	-0.19	0.47*	A-C* C-D**
	Parent Perception of School Services	-0.21*	--	-0.28*	0.17	A-D* C-D**
Teacher Report						
Head Start Year (0)						
Kindergarten (2)	Parent Participation	0.04	0.05	-0.12**	--	A-C** B-C**
1st grade (2)	Parent Participation	--	0.01	-0.14*	0.15	B-C* C-D**
3rd grade (2)	--	--	--	--	--	--

Key:
*** $p \leq 0.01$
** $p \leq 0.05$
* $p \leq 0.10$

Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.
Gray cell indicates no data available.
Note: + indicates scores from the Strengths and Difficulties Questionnaire that are reported as the percentage difference between Head Start and control group children in the normal category. The teacher-reported scores include Conduct Problems, Emotional Symptoms, Hyperactivity, Peer Problems, Pro-social Behavior, and Total Difficulties.

Exhibit E.6. Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Household Risk Index

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) Low/No Household Risk	Impact in Subgroup B (Head Start – Control) Moderate Household Risk	Impact in Subgroup C (Head Start – Control) High Household Risk	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE					
Direct Child Assessment					
Head Start Year (11)	Letter Naming	**2.10****	**4.49*****	0.24	A-B* B-C*
	Color Identification	**0.10*****	--	-0.08	A-C*
	WJ III Applied Problems	--	5.28**	-2.76	B-C*
Kindergarten (12)	--	--	--	--	--
1st grade (15)	PPVT (Adapted)	1.04	11.73**	--	A-B**
3rd grade (5)	--	--	--	--	--
School Performance					
Head Start Year (0)					
Kindergarten (5)	--	--	--	--	--
1st grade (5)	--	--	--	--	--
3rd grade (5)	Reading/Language Arts Ability	0.01	--	0.30*	A-C*
	Reading/Language Arts Skills	-0.09	--	0.69*	A-C**
SOCIAL-EMOTIONAL					
Parent Report					
Head Start Year (9)	Closeness	0.49**	-0.37	-0.55	A-B* A-C**
	Social Skills and Positive Approaches to Learning	-0.08	0.26	-0.66**	A-C* B-C**
	Social Competencies	0.00	--	-0.40*	A-C*
Kindergarten (9)	Positive Relationships	0.46	--	-2.66*	A-C*
1st grade (9)	--	--	--	--	--
3rd grade (5)	--	--	--	--	--

Exhibit E.6. Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Household Risk Index (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) Low/No Household Risk	Impact in Subgroup B (Head Start – Control) Moderate Household Risk	Impact in Subgroup C (Head Start – Control) High Household Risk	Statistically Significant Differences in Impacts Between Subgroups
Teacher Report					
Head Start Year (0)					
Kindergarten (11)	ASPI-Inattentive/Hyperactive	0.14	-3.93*	--	A-B*
	ASPI-Withdrawn/Low Energy	-0.41	2.45*	--	A-B**
	Closeness	0.07	--	3.41**	A-C**
	ASPI-Aggressive	0.63	--	-5.69**	A-C**
1st grade (11)	ASPI-Inattentive/Hyperactive	0.10	--	-3.50*	A-C*
	ASPI-Withdrawn/Low Energy	1.35**	-1.46	--	A-B*
3rd grade (10)	--	--	--	--	--
Child Report					
3rd grade (4)	Peer Relations	-0.09	-0.49**	0.35	A-B* B-C**
HEALTH					
Head Start Year (5)	Child Received Dental Care	**0.16*** **	0.01	**0.36*** **	A-B** A-C* **B-C*** **
	Child's Overall Health Status Is Excellent/Good	**-0.06****	--	0.18	A-C*
Kindergarten (5)	--	--	--	--	--
1st grade (5)	Child Received Dental Care	0.04	-0.13*	0.14	A-B** B-C*
	Child Had Care for Injury Last Month	0.01	0.01	**0.20*** **	**A-C**** B-C*
3rd grade (5)	Child Has Health Insurance Coverage	0.01	**-0.08****	--	A-B*

Exhibit E.6. Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Household Risk Index (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) Low/No Household Risk	Impact in Subgroup B (Head Start – Control) Moderate Household Risk	Impact in Subgroup C (Head Start – Control) High Household Risk	Statistically Significant Differences in Impacts Between Subgroups
PARENTING					
Parent Report					
Head Start Year (5)	--	--	--	--	--
Kindergarten (9)	Parental Cultural Enrichment Scale	0.24**	--	-0.31	A-C*
	Parenting Style: Authoritarian	-0.01	--	0.10*	A-C*
1st grade (8)	Parenting Style: Authoritative	0.08**	-0.05	--	A-B*
	Parenting Style: Permissive	-0.03	0.13**	--	A-B**
	Parenting Style: Neglectful	--	-0.09*	0.09	B-C**
3rd grade (10)	Doing Things Together	0.03	-0.01	-0.62**	A-C**
					B-C**
Teacher Report					
Head Start Year (0)					
Kindergarten (2)	--	--	--	--	--
1st grade (2)	--	--	--	--	--
3rd grade (2)	--	--	--	--	--

Key:
******* $p \leq 0.01$
** $p \leq 0.05$
* $p \leq 0.10$
Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.
Gray cell indicates no data available.

Exhibit E.7. Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Urbanicity

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) Not Urban	Impact in Subgroup B (Head Start –Control) Urban	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE				
Direct Child Assessment				
Head Start Year (11)	Letter Naming Counting Bears	**5.20***** -0.12	**1.82**** 0.07**	A-B** A-B*
Kindergarten (12)	WJ III Spelling	-7.86**	1.07	A-B**
1st grade (15)	--	--	--	--
3rd grade (5)	--	--	--	--
School Performance				
Head Start Year (0)				
Kindergarten (5)	--	--	--	--
1st grade (5)	Math Ability	-0.14**	-0.03	A-B*
3rd grade (5)	--	--	--	--
SOCIAL-EMOTIONAL				
Parent Report				
Head Start Year (9)	--	--	--	--
Kindergarten (9)	Total Problem Behavior	0.90**	-0.09	A-B**
1st grade (9)	--	--	--	--
3rd grade (5)	--	--	--	--
Teacher Report				
Head Start Year (0)				
Kindergarten (11)	--	--	--	--
1st grade (11)	--	--	--	--
3rd grade (10)	Emotional Symptoms+ Closeness with Teacher Positive Relationship with Teacher	0.01 **-2.42***** **-4.28*****	**-0.07***** -0.29 -0.73	A-B** A-B** A-B*
Child Report				
3rd grade (4)	Externalizing Peer Relations	**-0.45***** **-0.40****	-0.03 -0.09	A-B** A-B*

Exhibit E.7. Significant Subgroup Impacts and Differences in Impact for the 4-Year-Old Cohort: Urbanicity (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) Not Urban	Impact in Subgroup B (Head Start –Control) Urban	Statistically Significant Differences in Impacts Between Subgroups
HEALTH				
Head Start Year (5)	Child Received Dental Care	**0.32*** **	**0.12*** **	A-B*
Kindergarten (5)	--	--	--	--
1st grade (5)	--	--	--	--
3rd grade (5)	Child Received Dental Care	0.09**	0.00	A-B*
PARENTING				
Parent Report				
Head Start Year (5)	Parent Spanked Child in Last Week	-0.12*	0.01	A-B**
Kindergarten (9)	--	--	--	--
1st grade (8)	--	--	--	--
3rd grade (10)	Effect of Parenting on Parent's Life	**0.38*** **	-0.01	A-B**
Teacher Report				
Head Start Year (0)				
Kindergarten (2)	Parent Participation	-0.11**	0.01	**A-B*** **
1st grade (2)	School Contact and Communication	-0.15*	0.00	A-B*
3rd grade (2)	--	--	--	--

Key:
*** $p \leq 0.01$
** $p \leq 0.05$
* $p \leq 0.10$

Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.
Gray cell indicates no data available.
Note: + indicates scores from the Strengths and Difficulties Questionnaire that are reported as the percentage difference between Head Start and control group children in the normal category. The teacher-reported scores include Conduct Problems, Emotional Symptoms, Hyperactivity, Peer Problems, Pro-social Behavior, and Total Difficulties.

Exhibit E.8. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Child's Pre-Academic Skills

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) Children not in Lowest Quartile	Impact in Subgroup B (Head Start –Control) Children in Lowest Quartile	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE				
Direct Child Assessment				
Head Start Year (11)	--	--	--	--
Age 4 Year (11)	PPVT (Adapted)	-0.39	**8.87*** **	A-B**
	WJ III Applied Problems	-0.73	**5.03** **	A-B*
	WJ III Oral Comprehension	-0.50	**2.31** **	A-B*
	WJ III Pre-Academic Skills	0.02	**4.09** **	A-B*
	WJ III Letter Word Identification	1.07	**6.61** **	A-B*
Kindergarten (12)	--	--	--	--
1st grade (15)	--	--	--	--
3rd grade (5)	--	--	--	--
School Performance				
Head Start Year (0)				
Age 4 Year (0)				
Kindergarten (5)	--	--	--	--
1st grade (5)	--	--	--	--
3rd grade (5)	--	--	--	--
SOCIAL-EMOTIONAL				
Parent Report				
Head Start Year (9)	Withdrawn Behavior	0.02	-0.20*	A-B*
Age 4 Year (9)	Social Skills and Positive Approaches to Learning	0.06	**0.59*** **	A-B**
Kindergarten (9)	--	--	--	--
1st grade (9)	--	--	--	--
3rd grade (5)	Withdrawn Behavior	0.20**	-0.24	A-B**

Exhibit E.8. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Child's Pre-Academic Skills (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) Children not in Lowest Quartile	Impact in Subgroup B (Head Start –Control) Children in Lowest Quartile	Statistically Significant Differences in Impacts Between Subgroups
Teacher Report				
Head Start Year (0)				
Age 4 Year (0)				
Kindergarten (11)	ASPI-Aggressive ASPI-Problems with Peer Interaction	0.99* 1.56**	-1.49 -2.36	A-B* A-B**
1st grade (11)	--	--	--	--
3rd grade (10)	--	--	--	--
Child Report				
3rd grade (4)	--	--	--	--
HEALTH				
Head Start Year (5)	--	--	--	--
Age 4 Year (5)	--	--	--	--
Kindergarten (5)	--	--	--	--
1st grade (5)	--	--	--	--
3rd grade (5)	--	--	--	--
PARENTING				
Parent Report				
Head Start Year (5)	Parent Used Time Out in Last Week	-0.07**	0.04	A-B*
Age 4 Year (9)	Parent Read to Child in Last Week Parenting Style: Authoritarian	-0.04 -0.02	0.10* **-0.09*** **	A-B** A-B**
Kindergarten (9)	--	--	--	--
1st grade (8)	--	--	--	--
3rd grade (10)	--	--	--	--

Exhibit E.8. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Child's Pre-Academic Skills (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) Children not in Lowest Quartile	Impact in Subgroup B (Head Start –Control) Children in Lowest Quartile	Statistically Significant Differences in Impacts Between Subgroups
Teacher Report				
Head Start Year (0)				
Age 4 Year (0)				
Kindergarten (2)	--	--	--	--
1st grade (2)	--	--	--	--
3rd grade (2)	--	--	--	--

Key:
*** $p \leq 0.01$
** $p \leq 0.05$
* $p \leq 0.10$
Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.
Gray cell indicates no data available.

Exhibit E.9. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Child's Home Language

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) Dual Language Learners	Impact in Subgroup B (Head Start –Control) Children English Speaking	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE				
Direct Child Assessment				
Head Start Year (11)	PPVT (Adapted)	**11.87*** **	**4.97** **	A-B*
	WJ III Spelling	7.66**	0.62	A-B*
Age 4 Year (11)	CTOPPP Elision	**23.20*** **	4.00	**A-B** **
	Letter Naming	**3.20*** **	0.17	**A-B** **
	WJ III Pre-Academic Skills	**5.72** **	-0.07	**A-B** **
	WJ III Letter Word Identification	**8.59*** **	0.79	**A-B** **
	Counting Bears	**0.16*** **	-0.04	**A-B*** **
	WJ III Spelling	**4.75***	-1.03	**A-B***
Kindergarten (12)	WJ III Applied Problems	4.20	-2.56*	A-B**
	WJ III Pre-Academic Skills	5.20*	-1.68	A-B**
	WJ III Spelling	5.20*	-1.07	A-B*
	WJ III Math Reasoning	2.78	-2.07*	A-B**
1st grade (15)	--	--	--	--
3rd grade (5)	--	--	--	--
School Performance				
Head Start Year (0)				
Age 4 Year (0)				
Kindergarten (5)	--	--	--	--
1st grade (5)	--	--	--	--
3rd grade (5)	--	--	--	--
SOCIAL-EMOTIONAL				
Parent Report				
Head Start Year (9)	Withdrawn Behavior	0.19	-0.10*	A-B*
Age 4 Year (9)	Withdrawn Behavior	-0.24*	-0.03	A-B*
	Closeness	-0.77	0.54**	A-B**
Kindergarten (9)	--	--	--	--
1st grade (9)	--	--	--	--
3rd grade (5)	Social Skills and Positive Approaches to Learning	-0.06	0.31***	A-B*

Exhibit E.9. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Child's Home Language (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) Dual Language Learners	Impact in Subgroup B (Head Start –Control) Children English Speaking	Statistically Significant Differences in Impacts Between Subgroups
Teacher Report				
Head Start Year (0)				
Age 4 Year (0)				
Kindergarten (11)	--	--	--	--
1st grade (11)	--	--	--	--
3rd grade (10)	--	--	--	--
Child Report				
3rd grade (4)	--	--	--	--
HEALTH				
Head Start Year (5)	Child Received Dental Care Child's Overall Health Status Is Excellent/Good	**0.24***** **0.13*****	**0.14***** 0.02	A-B* A-B*
Age 4 Year (5)	--	--	--	--
Kindergarten (5)	Child Had Care for Injury Last Month	0.05*	-0.02	A-B*
1st grade (5)	--	--	--	--
3rd grade (5)	Child's Overall Health is Excellent/Good	**0.17*****	**-0.06****	**A-B*****
PARENTING				
Parent Report				
Head Start Year (5)	Parental Safety Practices Scale	-0.02	**0.05***	A-B*
Age 4 Year (9)	--	--	--	--
Kindergarten (9)	--	--	--	--
1st grade (8)	Parenting Style: Permissive	-0.07	0.05**	A-B*
3rd grade (10)	--	--	--	--

Exhibit E.9. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Child's Home Language (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) Dual Language Learners	Impact in Subgroup B (Head Start –Control) Children English Speaking	Statistically Significant Differences in Impacts Between Subgroups
Teacher Report				
Head Start Year (0)				
Age 4 Year (0)				
Kindergarten (2)	--	--	--	--
1st grade (2)	School Contact and Communication	**0.15*** **	-0.02	**A-B*** **
3rd grade (2)	--	--	--	--

Key:
******* $p \leq 0.01$
** $p \leq 0.05$
* $p \leq 0.10$
Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.
Gray cell indicates no data available.

Exhibit E.10. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Special Needs

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) Children With No Special Needs	Impact in Subgroup B (Head Start –Control) Children With Special Needs	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE				
Direct Child Assessment				
Head Start Year (11)	--	--	--	--
Age 4 Year (11)	--	--	--	--
Kindergarten (12)	--	--	--	--
1st grade (15)	WJ III Applied Problems	0.66	9.11*	A-B*
	WJ III Calculation	-0.70	5.86*	A-B*
	WJ III Quantitative Concept	0.08	6.64*	A-B*
	WJ III Math Reasoning	0.39	7.88**	A-B*
3rd grade (5)	--	--	--	--
School Performance				
Head Start Year (0)				
Age 4 Year (0)				
Kindergarten (5)	--	--	--	--
1st grade (5)	--	--	--	--
3rd grade (5)	--	--	--	--
Social-Emotional				
Parent Report				
Head Start Year (9)	--	--	--	--
Age 4 Year (9)	--	--	--	--
Kindergarten (9)	--	--	--	--
1st grade (9)	--	--	--	--
3rd grade (5)	--	--	--	--
Teacher Report				
Head Start Year (0)				
Age 4 Year (0)				
Kindergarten (11)	--	--	--	--

Exhibit E.10. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Special Needs (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) Children With No Special Needs	Impact in Subgroup B (Head Start –Control) Children With Special Needs	Statistically Significant Differences in Impacts Between Subgroups
1st grade (11)	ASPI-Inattentive/Hyperactive	-0.01	-3.84*	A-B*
	Conflict	0.23	-3.10*	A-B**
	Positive Relationships	0.05	3.56*	A-B*
	ASPI- Problems with Structured Learning	0.54	-5.01**	A-B**
3rd grade (10)	--	--	--	--
Child Report				
3rd grade (4)	--	--	--	--
HEALTH				
Head Start Year (5)	--	--	--	--
Age 4 Year (5)	Child Needs Ongoing Care	-0.01	0.15*	A-B**
Kindergarten (5)	Child Had Care for Injury Last Month	-0.01	**0.10*** **	**A-B*** **
1st grade (5)	--	--	--	--
3rd grade (5)	--	--	--	--
PARENTING				
Parent Report				
Head Start Year (5)	Parent Used Time Out in Last Week	-0.03	-0.15**	A-B*
Age 4 Year (9)	Parent Read to Child in Last Week	0.02	-0.17**	A-B**
Kindergarten (9)	Parent Spanked Child in Last Week	-0.03	-0.17**	A-B*
	Parenting Style: Authoritarian	-0.01	0.09**	A-B**
	Parenting Style: Authoritative	0.01	-0.12*	A-B*
1st grade (8)	Parent Used Time Out in Last Week	-0.02	**-0.30*** **	**A-B*** **
3rd grade (10)	--	--	--	--

Exhibit E.10. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Special Needs (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) Children With No Special Needs	Impact in Subgroup B (Head Start –Control) Children With Special Needs	Statistically Significant Differences in Impacts Between Subgroups
Teacher Report				
Head Start Year (0)				
Age 4 Year (0)				
Kindergarten (2)	--	--	--	--
1st grade (2)	--	--	--	--
3rd grade (2)	--	--	--	--

Key:
*** $p \leq 0.01$
** $p \leq 0.05$
* $p \leq 0.10$

Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.
Gray cell indicates no data available.

Exhibit E.11. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Biological Mother/ Caregiver Race/Ethnicity

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) White/Other Children	Impact in Subgroup B (Head Start – Control) Black Children	Impact in Subgroup C (Head Start – Control) Hispanic Children	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE					
Direct Child Assessment					
Head Start Year (11)	WJ III Oral Comprehension	2.41**	-2.07	--	A-B**
	WJ III Spelling	0.26	0.19	**6.70****	A-C *
					B-C**
	CTOPPP Elision	--	7.20*	-2.40	B-C*
Age 4 Year (11)	PPVT (Adapted)	--	-1.94	5.26*	B-C*
	WJ III Spelling	-2.40	--	3.64*	A-C*
	McCarthy Draw-a-Design	-0.38*	--	0.32	A-C**
	Counting Bears	--	-0.07	0.09**	B-C**
Kindergarten (12)	Letter Naming	0.38	--	-1.15**	A-C**
	WJ III Oral Comprehension	3.19**	-1.31	-0.34	A-B**
					A-C*
	WJ III Word Attack	3.39	-6.76*	--	A-B**
	WJ III Applied Problems	--	-3.86*	2.94	B-C**
1st grade (15)	PPVT (Adapted)	6.37**	-0.47	0.97	A-B*
					A-C*
	WJ III Quantitative Concept	--	-0.84	3.18*	B-C*
3rd grade (5)	--	--	--	--	--
School Performance					
Head Start Year (0)					
Age 4 Year (0)					
Kindergarten (5)	--	--	--	--	--
1st grade (5)	--	--	--	--	--
3rd grade (5)	Promotion	-0.04**	0.02	--	A-B**

Exhibit E.11. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Biological Mother/ Caregiver Race/Ethnicity (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) White/Other Children	Impact in Subgroup B (Head Start – Control) Black Children	Impact in Subgroup C (Head Start – Control) Hispanic Children	Statistically Significant Differences in Impacts Between Subgroups
SOCIAL-EMOTIONAL					
Parent Report					
Head Start Year (9)	Social Competencies	0.20*	-0.24*	--	A-B**
	Social Skills and Positive Approaches to Learning	0.30**	--	-0.19	A-C**
	Closeness	0.48*	--	-0.28	A-C*
Age 4 Year (9)	Closeness	0.68**	--	-0.35	A-C*
Kindergarten (9)	Closeness	0.60**	--	-0.12	A-C*
1st grade (9)	Conflict	0.47	-1.64*	--	A-B*
	Aggressive Behavior	0.40**	-0.33*	-0.22	**A-B*** **
					A-C**
	Total Problem Behavior	0.77*	-0.76**	-0.46	A-B**
					A-C**
	Positive Relationships	-0.50	2.18**	--	A-B**
	Hyperactive Behavior	0.18	--	-0.31**	A-C**
	Social Skills and Positive Approaches to Learning	0.34**	--	-0.23**	**A-C*** **
3rd grade (5)	Aggressive Behavior	0.31*	--	-0.24	A-C*
	Social Skills and Positive Approaches to Learning	0.60**	--	-0.12	A-C**
Teacher Report					
Head Start Year (0)					
Age 4 Year (0)					
Kindergarten (11)	ASPI-Inattentive/Hyperactive	1.07	0.61	-1.99**	A-C**
					B-C*
1st grade (11)	ASPI-Withdrawn/Low Energy	-0.62	--	1.56**	A-C**
3rd grade (10)	Closeness with Teacher	0.64	-1.22*	--	A-B*
Child Report					
3rd grade (4)	School	--	-0.17	0.22*	B-C**

Exhibit E.11. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Biological Mother/ Caregiver Race/Ethnicity (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) White/Other Children	Impact in Subgroup B (Head Start – Control) Black Children	Impact in Subgroup C (Head Start – Control) Hispanic Children	Statistically Significant Differences in Impacts Between Subgroups
HEALTH					
Head Start Year (5)	Child Had Care for Injury Last Month	**0.07*** **	-0.03	**-0.07***	**A-B** ** **A-C*** **
	Child Received Dental Care	**0.22*** **	0.09*	--	A-B*
	Child's Overall Health Status Is Excellent/Good	0.02	--	**0.12*** **	A-C*
Age 4 Year (5)	--	--	--	--	--
Kindergarten (5)	Child Has Health Insurance Coverage	-0.01	**0.08*** **	--	**A-B** **
	Child Needs Ongoing Care	**-0.10*** **	--	0.05	**A-C** **
1st grade (5)	Child Needs Ongoing Care	**-0.12*** **	0.05	0.04	**A-B*** ** **A-C*** **
3rd grade (5)	--	--	--	--	--
PARENTING					
Parent Report					
Head Start Year (9)	--	--	--	--	--
Age 4 Year (5)	--	--	--	--	--
Kindergarten (9)	Parenting Style: Authoritative	0.09*	-0.06	-0.03	A-B** A-C*
	Parental Cultural Enrichment Scale	-0.04	-0.18	0.22*	A-C* B-C**
	Parent Spanked Child in Last Week	-0.00	--	**-0.08** **	A-C*
	Parenting Style: Permissive	-0.03	0.07*	--	A-B*
	Parent Used Time Out in Last Week	-0.00	**-0.12*** **	--	A-B**
	Parental Safety Practices Scale	0.01	**-0.07*** **	**0.09** **	A-B* **B-C*** **

Exhibit E.11. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Biological Mother/ Caregiver Race/Ethnicity (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) White/Other Children	Impact in Subgroup B (Head Start – Control) Black Children	Impact in Subgroup C (Head Start – Control) Hispanic Children	Statistically Significant Differences in Impacts Between Subgroups
HEALTH					
Head Start Year (5)	Child Had Care for Injury Last Month	**0.07*** **	-0.03	**-0.07***	**A-B** ** **A-C*** **
	Child Received Dental Care	**0.22*** **	0.09*	--	A-B*
	Child's Overall Health Status Is Excellent/Good	0.02	--	**0.12*** **	A-C*
Age 4 Year (5)	--	--	--	--	--
Kindergarten (5)	Child Has Health Insurance Coverage	-0.01	**0.08*** **	--	**A-B** **
	Child Needs Ongoing Care	**-0.10*** **	--	0.05	**A-C** **
1st grade (5)	Child Needs Ongoing Care	**-0.12*** **	0.05	0.04	**A-B*** ** **A-C*** **
3rd grade (5)	--	--	--	--	--
PARENTING					
Parent Report					
Head Start Year (9)	--	--	--	--	--
Age 4 Year (5)	--	--	--	--	--
Kindergarten (9)	Parenting Style: Authoritative	0.09*	-0.06	-0.03	A-B** A-C*
	Parental Cultural Enrichment Scale	-0.04	-0.18	0.22*	A-C* B-C**
	Parent Spanked Child in Last Week	-0.00	--	**-0.08** **	A-C*
	Parenting Style: Permissive	-0.03	0.07*	--	A-B*
	Parent Used Time Out in Last Week	-0.00	**-0.12*** **	--	A-B**
	Parental Safety Practices Scale	0.01	**-0.07*** **	**0.09** **	A-B* **B-C*** **

Exhibit E.11. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Biological Mother/ Caregiver Race/Ethnicity (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) White/Other Children	Impact in Subgroup B (Head Start – Control) Black Children	Impact in Subgroup C (Head Start – Control) Hispanic Children	Statistically Significant Differences in Impacts Between Subgroups
1st grade (8)	Parent Spanked Child in Last Week	0.05	-0.07**	--	A-B**
	Parenting Style: Authoritative	--	-0.08*	0.05	B-C**
3rd grade (10)	Parent Spanked Child in Last Week	0.09*	--	-0.10	A-C**
Teacher Report					
Head Start Year (0)					
Age 4 Year (0)					
Kindergarten (2)	--	--	--	--	--
1st grade (2)	School Contact and Communication	-0.02	-0.02	0.11**	A-C* B-C*
3rd grade (2)	Parent Participation	0.06**	-0.07	-0.07**	A-B** **A-C***

Key:
*** $p \leq 0.01$
** $p \leq 0.05$
* $p \leq 0.10$
Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.
Gray cell indicates no data available.

Exhibit E.12. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Parent/Caregiver Reported Depressive Symptoms

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) No Symptoms	Impact in Subgroup B (Head Start – Control) Mild Symptoms	Impact in Subgroup C (Head Start – Control) Moderate Symptoms	Impact in Subgroup D (Head Start – Control) Severe Symptoms	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE						
Direct Child Assessment						
Head Start Year (11)	CTOPPP Elision	**9.01****	--	-5.37	--	A-C*
	Letter Naming	1.00	--	--	**2.76*****	A-D*
Age 4 Year (11)	CTOPPP Elision	**9.93*****	11.34*	-2.55	12.54*	**A-C**** B-C* C-D*
	WJ III Applied Problems	3.15*	--	-2.84	--	A-C*
	WJ III Letter Word Identification	5.23**	--	-5.70	--	**A-C*****
	WJ III Pre-Academic Skills	3.52**	1.85	-5.51*	--	**A-C**** B-C*
	WJ III Spelling	3.12	0.06	**-10.81*****	--	**A-C***** B-C**
	Letter Naming	1.14*	--	-1.96	--	**A-C***
	McCarthy Draw-a-Design	0.00	0.24	-0.52**	--	A-C* B-C**
	PPVT (Adapted)	**7.38*****	1.16	-9.81**	--	**A-C***** B-C*
Kindergarten (12)	WJ III Letter Word Identification	4.73**	-4.28	--	--	**A-B*****
1st grade (15)	PPVT (Adapted)	**6.47****	-0.46	-6.48	--	A-B* **A-C****
	WJ III Pre-Academic Skills	3.96*	-3.80	-7.53*	5.81	A-B* **A-C***** C-D*

Exhibit E.12. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Parent/Caregiver Reported Depressive Symptoms (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) No Symptoms	Impact in Subgroup B (Head Start – Control) Mild Symptoms	Impact in Subgroup C (Head Start – Control) Moderate Symptoms	Impact in Subgroup D (Head Start – Control) Severe Symptoms	Statistically Significant Differences in Impacts Between Subgroups
1st grade (15)	WJ III Basic Reading Skills	4.98*	-5.93	-10.03*	8.82	A-B* **A-C***** C-D*
	WJ III Applied Problems	**4.16****	-3.37	-2.07	--	A-B ** **A-C***
	WJ III Letter Word Identification	6.30*	-6.09	-13.61**	6.87	A-B* **A-C***** C-D*
	WJ III Oral Comprehension	**3.45****	1.65	-5.15***	1.85	**A-C**** B-C** C-D*
	WJ III Academic Applications	**3.45****	--	-5.06**	3.98	**A-C***** C-D*
	WJ III Math Reasoning	**3.22****	--	-3.15	--	**A-C****
	WJ III Academic Skills	2.27	--	-8.45**	5.46	**A-C***** C-D*
	WJ III Writing Sample	1.53	--	-4.43**	--	**A-C*****
	WJ III Passage Comprehension	**4.89****	--	-9.26**	2.15	**A-C***** C-D*
	WJ III Spelling	1.76	--	-8.16*	--	**A-C*****
3rd grade (5)	PPVT	4.48**	--	-4.05	--	A-C*
	ECLS-K Reading	3.13**	-2.59		--	A-B*
	WJIII Letter-Word Identification	5.79*	-3.89	-7.48	--	A-B** A-C**
School Performance						
Head Start Year (0)						
Age 4 Year (0)						
Kindergarten (5)	Language and Literacy Ability	--	0.06	**-0.16****	--	B-C*
	Math Ability	--	0.02	**-0.19****	-0.19*	B-C** B-D*

Exhibit E.12. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Parent/Caregiver Reported Depressive Symptoms (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) No Symptoms	Impact in Subgroup B (Head Start – Control) Mild Symptoms	Impact in Subgroup C (Head Start – Control) Moderate Symptoms	Impact in Subgroup D (Head Start – Control) Severe Symptoms	Statistically Significant Differences in Impacts Between Subgroups
1st grade (5)	Language and Literacy Ability	0.03	-0.11*	--	--	A-B**
	Math Ability	-0.01	**-0.11****	--	--	A-B*
	Promotion	0.02	--	**-0.11****	0.06	**A-C***** **C-D****
	Science and Social Studies Ability	--	**-0.11****	--	0.11	B-D*
3rd grade (5)	--	--	--	--	--	--
SOCIAL-EMOTIONAL						
Parent Report						
Head Start Year (9)	Hyperactive Behavior	-0.33**	-0.46**	0.27	-0.77*	A-C* B-C** C-D**
Age 4 Year (9)	Hyperactive Behavior	**-0.37*****	-0.03	0.39**	0.27	A-B* **A-C***** A-D* B-C*
	Total Problem Behavior	-0.61**	--	0.38	--	A-C*
	Closeness	0.26	0.13	--	1.57**	A-D* B-D*
Kindergarten (9)	Closeness	0.83**	-0.65**	-0.90**	--	A-B*** A-C***
	Social Competencies	0.22*	-0.10	--	-0.39	A-B * A-D*
	Social Skills and Positive Approaches to Learning	**0.46*****	-0.06	--	--	A-B **
	Hyperactive Behavior	-0.31**	--	0.34	--	A-C**
	Aggressive Behavior	-0.27**	--	--	0.68	A-D **
1st grade (9)	--	--	--	--	--	--
3rd grade (5)	--	--	--	--	--	--

Exhibit E.12. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Parent/Caregiver Reported Depressive Symptoms (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) No Symptoms	Impact in Subgroup B (Head Start – Control) Mild Symptoms	Impact in Subgroup C (Head Start – Control) Moderate Symptoms	Impact in Subgroup D (Head Start – Control) Severe Symptoms	Statistically Significant Differences in Impacts Between Subgroups
Teacher Report						
Head Start Year (0)						
Age 4 Year (0)						
Kindergarten (11)	ASPI-Problem with Peer Interaction	0.08	2.45*	5.01**	-2.89	A-B* A-C** C-D**
	Conflict	--	1.82*	1.97*	-2.84	B-D* C-D**
	Positive Relationships	--	-2.55**	-3.38**	3.65	B-D* C-D**
	ASPI-Aggressive	-0.28	--	3.23**	-1.69	A-C** C-D*
	ASPI-Oppositional	--	--	2.44*	-3.34	C-D**
1st grade (11)	ASPI-Withdrawn/Low Energy	0.07	1.98**	--	--	A-B*
	ASPI-Shy/Social Reticent	0.39	**2.36*** **	--	-2.18	A-B* B-D**
	Positive Relationships	1.11*	-1.55	-3.32*	5.27*	A-B* A-C** B-D* C-D**
	Conflict	-0.53	0.90	2.79**	-4.19*	A-C** A-D* B-D** C-D**
	ASPI- Problems with Structured Learner	0.01	**3.08*** **	-1.39	-3.30	A-B** B-C* B-D**

Exhibit E.12. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Parent/Caregiver Reported Depressive Symptoms (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) No Symptoms	Impact in Subgroup B (Head Start – Control) Mild Symptoms	Impact in Subgroup C (Head Start – Control) Moderate Symptoms	Impact in Subgroup D (Head Start – Control) Severe Symptoms	Statistically Significant Differences in Impacts Between Subgroups
3rd grade (10)	Emotional Symptoms +	0.03	0.07	-0.01	-0.15**	A-D*** B-D*** C-D**
	Pro-social Behavior	--	**-0.18*** **	0.08	--	B-C*
	Positive Relationship with Teacher	-0.32	**-3.20*** **	--	--	A-B*
Child Report						
3rd grade (4)	Peer Relations	0.05	-0.07	0.58***	0.01	A-C** B-C** C-D**
HEALTH						
Head Start Year (5)	Child Has Health Insurance Coverage	--	--	-0.03	0.12*	C-D**
	Child Received Dental Care	--	--	0.08	**0.28*** **	C-D*
Age 4 Year (5)	Child's Overall Health Status Is Excellent/Good	0.04	-0.02	**-0.16*** **	0.11	**A-C*** ** B-C* C-D**
	Child Has Health Insurance Coverage	0.03	0.03	**-0.06** **	--	**A-C** ** B-C*
Kindergarten (5)	--	--	--	--	--	--
1st grade (5)	Child Had Care for Injury in Last Month	0.01	-0.01	**0.11** **	--	A-C* **B-C** **
	Child Needs Ongoing Care	--	0.06	--	-0.15*	B-D**
3rd grade (5)	Child Had Care for Injury Last Month	-0.02	0.00	-0.15**	--	A-C* B-C*

Exhibit E.12. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Parent/Caregiver Reported Depressive Symptoms (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) No Symptoms	Impact in Subgroup B (Head Start – Control) Mild Symptoms	Impact in Subgroup C (Head Start – Control) Moderate Symptoms	Impact in Subgroup D (Head Start – Control) Severe Symptoms	Statistically Significant Differences in Impacts Between Subgroups
PARENTING						
Parent Report						
Head Start Year (5)	Parent Read to Child in Last Week	--	--	**0.16*** **	-0.03	C-D*
	Parental Safety Practices Scale	**0.08*** **	-0.02	-0.06	--	**A-B** ** **A-C*** **
Age 4 Year (9)	Parental Safety Practices Scale	0.06**	-0.04	--	--	A-B*
Kindergarten (9)	Parent Read to Child in Last Week	-0.03	--	0.13*	--	A-C*
	Parental Safety Practices Scale	-0.01	0.09**	--	--	A-B**
	Parenting Style: Neglect	-0.05**	--	--	0.09	A-D**
	Parenting Style: Permissive	--	-0.08	0.10*	--	B-C**
	Parent Used Time Out in Last Week	--	--	-0.17**	0.14	C-D**
1st grade (8)	Parent Spanked Child in Last Week	-0.08**	--	--	0.10	A-D**
3rd grade (10)	Supportive School Environment	-0.01	-0.03	-0.08	0.35**	A-D** B-D* C-D**
	Parenting Style: Authoritarian	-0.01	0.01	**-0.19*** **	0.04	A-C** B-C** C-D**
	Parenting Style: Authoritative	--	--	**0.27*** **	-0.01	C-D*
	Effect of Parenting on Parent's Life	-0.13	--	0.25*	--	A-C*
	Time Spent With Child	-0.22***	--	0.17	0.30	A-C* A-D**

Exhibit E.12. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Parent/Caregiver Reported Depressive Symptoms (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) No Symptoms	Impact in Subgroup B (Head Start – Control) Mild Symptoms	Impact in Subgroup C (Head Start – Control) Moderate Symptoms	Impact in Subgroup D (Head Start – Control) Severe Symptoms	Statistically Significant Differences in Impacts Between Subgroups
Teacher Report						
Head Start Year (0)						
Age 4 Year (0)						
Kindergarten (2)	School Contact and Communication	-0.00	0.06	**-0.13****	0.10	A-C** B-C* C-D**
1st grade (2)	--	--	--	--	--	--
3rd grade (2)	--	--	--	--	--	--

Key:
******* $p \leq 0.01$
** $p \leq 0.05$
***** $p \leq 0.10$

Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.
Gray cell indicates no data available.
Note: + indicates scores from the Strengths and Difficulties Questionnaire that are reported as the percentage difference between Head Start and control group children in the normal category. The teacher-reported scores include Conduct Problems, Emotional Symptoms, Hyperactivity, Peer Problems, Pro-social Behavior, and Total Difficulties.

Exhibit E.13. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Household Risk Index

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) Low/No Household Risk	Impact in Subgroup B (Head Start – Control) Moderate Household Risk	Impact in Subgroup C (Head Start – Control) High Household Risk	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE					
Direct Child Assessment					
Head Start Year (11)	Counting Bears	**0.06****	-0.07	--	A-B*
	CTOPPP Elision	**7.81****	-6.14	--	A-B**
	McCarthy Draw-a-Design	**0.16****	0.02	0.59**	A-C*
					B-C**
	WJ III Spelling	--	-1.67	8.84**	B-C**
Age 4 Year (11)	WJ III Letter Word Identification	4.28**	-3.00	--	A-B*
	CTOPPP Elision	7.19**	6.95	**25.78*****	A-C**
					B-C*
	WJ III Pre-Academic Skills	2.77*	-4.65	--	A-B**
Kindergarten (12)	WJ III Applied Problems	0.30	-5.08**	--	A-B*
	WJ III Math Reasoning	-0.16	-3.94**	--	A-B*
	Letter Naming	-0.04	-2.05**	1.84	A-B**
					B-C***
	WJ III Spelling	0.30	-1.68	8.57**	A-C*
					B-C*
1st grade (15)	WJ III Calculation	0.79	-4.71*	--	A-B**
	WJ III Spelling	-0.58	-8.56**	8.85**	A-B*
					A-C**
					B-C***
	WJ III Academic Skills	--	-5.46	7.26*	B-C**
	WJ III Passage Comprehension	--	-4.85	7.87**	B-C**
	WJ III Pre-Academic Skills	--	-4.01	7.62*	B-C**
	WJ III Academic Applications	--	-2.77	5.44*	B-C**
3rd grade (5)	ECLS-K Reading	-0.16	-3.75	**8.85*****	**A-C*****
					B-C***
	WJIII Letter-Word Identification	0.71	-6.55	**15.82*****	**A-C*****
					B-C***

Exhibit E.13. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Household Risk Index (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) Low/No Household Risk	Impact in Subgroup B (Head Start – Control) Moderate Household Risk	Impact in Subgroup C (Head Start – Control) High Household Risk	Statistically Significant Differences in Impacts Between Subgroups
School Performance					
Head Start Year (0)					
Age 4 Year (0)					
Kindergarten (5)	--	--	--	--	--
1st grade (5)	--	--	--	--	--
3rd grade (5)	Math Skills	0.05	-0.31**	0.26	A-B** B-C*
	Reading/Language Arts Skills	-0.05	-0.26	0.40*	A-C* **B-C***
SOCIAL-EMOTIONAL					
Parent Report					
Head Start Year (9)	Total Problem Behavior	--	-1.08**	0.38	B-C*
Age 4 Year (9)	Aggressive Behavior	-0.04	-0.50*	--	A-B*
Kindergarten (9)	--	--	--	--	--
1st grade (9)	Aggressive Behavior	0.04	-0.57**	0.15	A-B** B-C*
	Hyperactive Behavior	-0.01	**-0.50*** **	--	A-B**
	Total Problems Behavior	0.16	**-1.59*** **	--	**A-B*** **
3rd grade (5)	--	--	--	--	--
Teacher Report					
Head Start Year (0)					
Age 4 Year (0)					
Kindergarten (11)	Conflict	-0.40	--	3.05*	A-C*
	Positive Relationship	-0.04	--	-4.06*	A-C*
1st grade (11)	ASPI—Problems with Peer Interaction	0.46	-4.09**	--	A-B**
3rd grade (10)	Pro-social Behavior+	-0.08**	0.06	--	A-B**
Child Report					
3rd grade (4)	--	--	--	--	--

Exhibit E.13. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Household Risk Index (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) Low/No Household Risk	Impact in Subgroup B (Head Start – Control) Moderate Household Risk	Impact in Subgroup C (Head Start – Control) High Household Risk	Statistically Significant Differences in Impacts Between Subgroups
HEALTH					
Head Start Year (5)	Child Received Dental Care	**0.19*** **	0.01	**0.25** **	**A-B ** ** **B-C** **
	Child Had Care for Injury in Last Month	0.01	**-0.11** **	0.09	**A-B** ** **B-C** **
Age 4 Year (5)	Child Had Care for Injury in Last Month	0.02	--	0.12**	A-C*
	Child Needs Ongoing Care	-0.01	0.09*	-0.07	A-B* B-C*
Kindergarten (5)	--	--	--	--	--
1st grade (5)	--	--	--	--	--
3rd grade (5)	Child Needs Ongoing Care	-0.01	**0.18*** **	-0.06	**A-B*** ** B-C**
PARENTING					
Parent Report					
Head Start Year (5)	--	--	--	--	--
Age 4 Year (9)	--	--	--	--	--
Kindergarten (9)	--	--	--	--	--
1st grade (8)	Parent Spanked Child in Last	-0.01	-0.15***	0.12	A-B** **B-C*** **
3rd grade (10)	Doing Things Together	--	-0.11	0.36**	B-C *
	Parenting Style: Authoritarian	0.01	-0.14**	--	**A-B** **
	Parenting Style; Authoritative	0.07*	--	0.26**	A-C*
	Parenting Style: Permissive	-0.06*	0.11*	-0.19*	**A-B** ** B-C **
	Effect of Parenting on Parent's Life	-0.11*	0.21	--	A-B**

Exhibit E.13. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Household Risk Index (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) Low/No Household Risk	Impact in Subgroup B (Head Start – Control) Moderate Household Risk	Impact in Subgroup C (Head Start – Control) High Household Risk	Statistically Significant Differences in Impacts Between Subgroups
Teacher Report					
Head Start Year (0)					
Age 4 Year (0)					
Kindergarten (2)	--	--	--	--	--
1st grade (2)	--	--	--	--	--
3rd grade (2)	--	--	--	--	--

Key:
******* $p \leq 0.01$
** $p \leq 0.05$
* $p \leq 0.10$
Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.
Gray cell indicates no data available.
Note: + indicates scores from the Strengths and Difficulties Questionnaire that are reported as the percentage difference between Head Start and control group children in the normal category. The teacher-reported scores include Conduct Problems, Emotional Symptoms, Hyperactivity, Peer Problems, Pro-social Behavior, and Total Difficulties.

Exhibit E.14. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Urbanicity

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) Not Urban	Impact in Subgroup B (Head Start –Control) Urban	Statistically Significant Differences in Impacts Between Subgroups
COGNITIVE				
Direct Child Assessment				
Head Start Year (11)	WJ III Pre-Academic Skills	**8.37*** **	3.21**	A-B*
	McCarthy Draw-a-Design	**0.42*** **	0.09	A-B**
	Letter Naming	**3.11*** **	1.18**	A-B**
	WJ III Letter Word Identification	**11.75*** **	**4.72*** **	A-B**
Age 4 Year (11)	WJ III Applied Problems	5.21**	-0.34	A-B*
Kindergarten (12)	WJ III Spelling	6.05*	-0.94	A-B*
1st grade (15)	WJ III Pre-Academic Skills	7.06**	-1.45	A-B**
	WJ III Spelling	6.04*	-3.01	A-B**
	WJ III Academic Applications	5.00**	-0.34	A-B*
	WJ III Word Attack	10.08*	-3.29	A-B**
	WJ III Writing	4.67**	-1.28	A-B**
	WJ III Basic Reading	9.64*	-2.53	A-B**
	WJ III Math Reasoning	5.04**	0.26	A-B*
3rd grade (10)	--	--	--	--
School Performance				
Head Start Year (0)				
Age 4 Year (0)				
Kindergarten (5)	--	--	--	--
1st grade (5)	--	--	--	--
3rd grade (5)	--	--	--	--
SOCIAL-EMOTIONAL				
Parent Report				
Head Start Year (9)	Aggressive Behavior	**-0.40** **	-0.02	A-B**
	Social Competencies	**0.28** **	-0.11	**A-B*** **
	Total Problem Behavior	**-1.04*** **	-0.38**	A-B*
Age 4 Year (9)	Aggressive Behavior	-0.45**	-0.05	A-B*
Kindergarten (9)	--	--	--	--
1st grade (9)	--	--	--	--
3rd grade (5)	--	--	--	--

Exhibit E.14. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Urbanicity (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) Not Urban	Impact in Subgroup B (Head Start –Control) Urban	Statistically Significant Differences in Impacts Between Subgroups
Teacher Report				
Head Start Year (0)				
Age 4 Year (0)				
Kindergarten (11)	--	--	--	--
1st grade (11)	--	--	--	--
3rd grade (10)	Pro-social Behavior+ Total Difficulties+	-0.15** 0.11*	-0.02 -0.02	A-B* A-B*
Child Report				
3rd grade (4)	School	0.23**	0.00	A-B*
HEALTH				
Head Start Year (5)	Child Had Care for Injury in Last Month	**0.08****	-0.03*	**A-B*****
Age 4 Year (5)	Child's Overall Health Status Is Excellent/Good	**0.11****	-0.02	**A-B****
Kindergarten (5)	Child Needs Ongoing Care	**-0.11****	-0.01	A-B**
1st grade (5)	--	--	--	--
3rd grade (5)	--	--	--	--
PARENTING				
Parent Report				
Head Start Year (5)	Parental Safety Practices Scale	**0.09****	0.02	A-B *
Age 4 Year (9)	--	--	--	--
Kindergarten (9)	--	--	--	--
1st grade (8)	--	--	--	--
3rd grade (10)	Effect of Parenting on Parent's Life	-0.31**	0.03	A-B**

Exhibit E.14. Significant Subgroup Impacts and Differences in Impact for the 3-Year-Old Cohort: Urbanicity (continued)

Year (Number of Outcomes Examined)	Outcome Measure	Impact in Subgroup A (Head Start – Control) Not Urban	Impact in Subgroup B (Head Start –Control) Urban	Statistically Significant Differences in Impacts Between Subgroups
Teacher Report				
Head Start Year (0)				
Age 4 Year (0)				
Kindergarten (2)	--	--	--	--
1st grade (2)	--	--	--	--
3rd grade (2)	--	--	--	--

Key:
*** $p \leq 0.01$
** $p \leq 0.05$
* $p \leq 0.10$

Bold exhibit entry indicates that the impact passes the Benjamini-Hochberg test for multiple comparisons with a 10 percent false discovery rate.
Gray cell indicates no data available.
Note: + indicates scores from the Strengths and Difficulties Questionnaire that are reported as the percentage difference between Head Start and control group children in the normal category. The teacher-reported scores include Conduct Problems, Emotional Symptoms, Hyperactivity, Peer Problems, Pro-social Behavior, and Total Difficulties.