

**U.S. Department of Health and Human Services
Office of Planning, Research & Evaluation
Administration for Children and Families**

**Third Grade Follow-Up to the
Head Start Impact Study**

*Office of Management and Budget
Clearance Package Supporting Statement
And Data Collection Instruments*

November 3, 2006

TABLE OF CONTENTS

A. Justification..... 1

A.1 Explanation Of The Circumstances Which Make The Data Collection Necessary..... 1

A.2 How The Information Will Be Collected, By Whom It Will Be Collected, And For What Purpose..... 10

A.3 Use Of Automated, Electronic, Mechanical, Or Other Technological Collection Techniques..... 16

A.4 Efforts To Identify Duplication..... 18

A.5 Minimizing Impact On Small Businesses Or Other Small Entities..... 19

A.6 Consequences If The Collection Is Not Conducted..... 19

A.7 Special Circumstances..... 19

A.8 Consultation With Persons Outside The Agency..... 19

A.9 Remuneration To Respondents..... 20

A.10 Assurances Of Privacy..... 23

A.11 Questions Of A Sensitive Nature..... 23

A.12 Respondent Burden..... 24

A.13 Total Annual Cost Burden..... 27

A.14 Annualized Cost To The Government..... 27

A.15 Reasons For Any Program Changes..... 27

A.16 Plans For Tabulation And Statistical Analysis And Time Schedule..... 27

A.17 Approval To Not Display The OMB Expiration Date..... 46

A.18 Exception To The Certification Statement..... 46

B. Collections Of Information Employing Statistical Methods..... 49

B.1 Potential Respondent Universe..... 49

B.2 Description Of Sampling And Information Collection Procedures..... 49

B.3 Methods To Maximize Response..... 58

B.4 Tests Of Procedures To Minimize Burden..... 60

B.5 Identity Of Individuals Consulted On Statistical Aspects Of Design And Identity Of Contractors..... 61

Bibliography..... 62

Appendix A Data Collection Instruments

Appendix B Federal Register Announcement

Appendix C Public Comments

Appendix D Advisory Committee Members

Appendix E Head Start Impact Study Sampling Plan

LIST OF TABLES

Table 1	Number of Children in the Head Start and Non-Head Start Groups by Age Cohort.....	11
Table 2	Percent of HSIS Parent Interviews and Child Assessments Complete By Data Collection Period.....	11
Table 3	Data Collection Schedule.....	12
Table 4	Data Collection Instruments and Activities.....	22
Table 5a	Estimated Response Burden for Respondents in the Third Grade Follow-Up to the Head Start Impact Study-Spring 2007.....	25
Table 5b	Estimated Response Burden for Respondents in the Third Grade Follow-Up to the Head Start Impact Study-Spring 2008.....	26
Table 6	Hypothetical Example: Impact of Head Start on Access to “Quality” elementary School Teachers.....	29
Table 7	Project Deliverables and Delivery Dates.....	47
Table 8	Expected Sample Sizes in the Third Grade Follow-Up Study.....	50
Table 9a	Minimum Detectable Effect Sizes with Power = .8 for the Third Grade Follow-Up Sample.....	54
Table 9b	Minimum Detectable Differences in Proportions with Power = .8 for the Third Grade Follow-Up Sample, Based on Spring 2003 HSIS Data.....	54

LIST OF FIGURES

Figure 1	Unconditional and Conditional Latent Growth Models.....	43
----------	---	----

LIST OF INSTRUMENTS

1. Child Assessment (Spanish and Bilingual versions available)
2. Parent Interview (Spanish version available)
3. Teacher Survey and Teacher/Child Report Form (Spanish version available)
4. Principal Survey (Spanish version available)

A. JUSTIFICATION

A.1 Explanation of the Circumstances Which Make the Data Collection Necessary

The Office of the Administration for Children and Families, U.S. Department of Health and Human Services (DHHS), is submitting this *Request for OMB Review* in support of the Third Grade Follow-Up to the Head Start Impact Study (Third Grade Follow-Up Study). OMB approved the initial package for the Head Start Impact Study (HSIS) in September 2002 (OMB # 0970-0229, Expiration Date: 09/30/2005 and the HSIS continuation OMB # 0970-0029, Expiration Date: 07/30/2006). The purpose of the Third Grade Follow-Up Study is to build 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. Specifically, the study will examine the degree to which the impacts of Head Start on initial school readiness are altered, maintained or perhaps expanded by children's school experiences and the various school quality and family/community factors that come into play up to and during the third grade. The Third Grade Follow-Up Study also is designed to build on the comprehensive instrument design and data collection plans effectively implemented in the HSIS.

Background

Overview of Head Start. Over the years, Head Start has served nearly 23 million preschool children and their families since it began in 1965 as a six-week summer program for children of low-income families. The program provides comprehensive early child development services to low-income children, their families, and communities. Head Start has evolved over time to include a wide variety of program options based on the specific situations and resources of the communities to meet the changing needs of the children and families it serves. Variations in services include, but are not limited to, programs offering center-based services, home-based services, part-day enrollment, full-day enrollment and/or one or two years of services. In addition, many programs are now partnering with non-Head Start agencies and/or combining funds from various sources to coordinate services that best address the needs of children and families.

As Head Start's Federal appropriation has grown, (\$96 million in summer 1965 to \$6.8 billion in 2005) so have initiatives calling for improved outcomes and accountability (e.g., Chief

Financial Officers Act, Government Performance and Results Act of 1993). During the rapid expansion of Head Start, the U.S. General Accounting Office (GAO) released two reports underlining the lack of rigorous research on Head Start's effectiveness noting that "...the body of research on current Head Start is insufficient to draw conclusions about the impact of the national program" (GAO, 1997). The 1998 report added, "...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 upon GAO recommendation, and the testimony of research methodologists and early childhood experts, Congress mandated in Head Start's 1998 reauthorization that DHHS conduct research to determine, on a national level, the impact of Head Start on the children it serves. Congress called for an expert panel to develop recommendations regarding the study design to "...determine if, overall, the Head Start programs have impacts consistent with their primary goal of increasing the social competence of children, by increasing the every day effectiveness of the children in dealing with their present environments and future responsibilities, and increasing their school readiness" (42 USC 9801, et.seq.). The research should also consider variables such as whether Head Start strengthens families as the nurturers of their children and increases children's access to other education, health, nutritional, and community services.

To design such a study, the Department convened a committee of distinguished experts, the Advisory Committee on Head Start Research and Evaluation, that considered the major issues and challenges in designing a rigorous research study that is both credible and feasible, and the committee recommended a framework for the design of the Head Start Impact Study (HSIS). A contract was awarded in October 2000 to Westat, in collaboration with the Urban Institute, American Institutes for Research, and Decision Information Resources to conduct the Head Start Impact Study as mandated by the Coats Human Services Amendments of 1998, PL 105-285.

The National Head Start Impact Study is a longitudinal study that involved approximately 5,000 three- and four- year old preschool children across an estimated 75 nationally representative grantee/delegate agencies (in communities where there are more eligible children and families than can be served by the program). Children were randomly assigned to either a Head Start group that had access to Head Start program services or to a non-Head Start group that

could enroll in available community non-Head Start services, selected by their parents. Data collection began in fall 2002 and continued through spring 2006, following children through the spring of their first-grade year. The HSIS data collection included parent interviews, teacher and care provider surveys, child assessments, direct observations of the quality in different care settings, and teacher/care provider ratings of children. Although the Third Grade Follow-Up Study will be based largely on work that has already been completed for the HSIS, it offers an opportunity to examine the degree to which the impacts of Head Start on initial school readiness are changed by children's third grade school experiences and the family/community factors associated with the child during the school years. This study allows us to broaden the scope of analysis to include factors that have not yet been examined as part of the HSIS.

The implementation of the Third Grade Follow-Up Study must be understood within the history of the preschool and Head Start research and evaluation efforts. Crucial to the study is understanding the evidence as it concerns preschool experiences influence on outcomes in elementary school.

Preschool Intervention Studies. Unfortunately, there is a paucity of experimental design studies examining preschool intervention and even fewer preschool intervention studies with longitudinal designs that stretch beyond kindergarten or first grade. Use of experimental design is concentrated on a few studies such as the Abecedarian Project, Project CARE, and the Early Training Project. These studies randomize families matched on control variables (e.g., income, gender) and place some into preschool intervention while excluding others. This allows researchers to determine the effects of treatment by comparing treated children and families to those that were similar at the start of the study and whose experiences differ only in terms of whether or not they received the intervention. When the experiments involve high intensity programs (i.e. extensive instruction, comprehensive services, home visits), generalization is difficult. High intensity projects are often considered too costly and resource intensive to be replicated on a national scale. It is often their small sample size that makes them feasible for researchers to conduct them. For example, the initial Abecedarian sample consisted of 117 participants (Campbell & Ramey, 1995). Moreover, findings from intense programs cannot always be expected to be replicated by more moderate programs. This is due to evidence that the intensity of a preschool intervention can increase the positive effect those programs exert on child outcomes (Nelson et al., 2003; Ramey & Ramey, 2004).

More common are studies that explore early experience predictors to school-age outcomes (e.g., Miles & Stipek, 2006; Peisner-Feinberg et al., 2001) and quasi-experimental

designs intended to determine the causal contribution of specific experiences or programs to those outcomes. Quasi-experimental include wait-list designs that compare children who receive an intervention such as Head Start to those who are waiting for the opportunity to enroll, and regression-discontinuity approaches that rank children on level of need and adjust for these and other differences in comparing outcomes of participants and nonparticipants. The Chicago Longitudinal Study and the work of Abbott-Shim and colleagues at the Georgia State University Quality Research Center (2003) are two examples of such work. Although ethical concerns often call for them, quasi-experimental designs generally can not provide the clarity of data obtained through true experimental studies. Other studies that explore relationships between variables provide useful information to guide research hypotheses regarding the aspects of preschool interventions that are most likely to predict later child outcomes. For example, Peisner-Feinberg and colleagues (2001) reported a positive relationship between the quality of preschool care and elementary math scores. However, the question of long-term impacts of preschool interventions is best answered through longitudinal data from studies allowing the direct comparison of children who received the intervention to those who did not—and, ideally, comparing sets of children who are indistinguishable at the outset by virtue of having been selected at random from a common pool of eligible applicants.

Summarized below are some of the findings about the effects of preschool participation on children's later outcomes.

Cognitive Outcomes. It is clear that preschool participation can have lasting cognitive and academic effects (Barnett, 1995; Miller & Bizzell, 1984; Nelson, Westhues, & MacLeod, 2003). For example, children who attend preschool are less likely to be held back at grade level or to be in special education classes (Darlington, Royce, Snipper, Murray, & Lazar, 1980). Further, the NICHD Early Child Care Research Study linked high-quality child care with higher school-age math and reading test scores (NICHD ECCRN, 2005).

Social Outcomes. Social outcomes such as socialization skills (e.g., Barnett, 1995; Hubbs-Tait et al., 2002) and juvenile delinquency (e.g., Garces, Thomas, & Curie, 2002; Reynolds, Ou, & Topitzes, 2004) are also positively influenced by preschool attendance. For example, children observed to have close relationships with their preschool teacher have been found to have higher attention and sociability ratings in the second grade as well as displaying fewer problem behaviors (Peisner-Feinberg et al., 2001). Further, in their meta-analysis of over 60 studies, Paro and Pianta (2000) concluded that measures of social outcomes taken soon after

preschool explained a significant portion of the variance for assessments of social outcomes taken later in elementary school, albeit with small effect sizes.

Experimental Evidence. Experimental data on preschool provides further evidence of the long-term effects of preschool intervention. For example, the Abecedarian study reported that children in a preschool intervention group performed better on cognitive tests in third grade than those who had not had the intervention (Campbell & Ramey, 1995). Similar results were found for fourth graders who had summer interventions during the preschool period (Gray & Klaus, 1970); children who had received intervention outperformed control children on intelligence tests. It should be noted that the entire sample in the Campbell and Ramey (1995) study saw a decline in cognitive scores following the first grade, however, children who had intensive preschool interventions experienced less change over time.

Head Start Evidence Studies examining the effects of a national program such as Head Start have the potential to be more generalizable. However, taken as a whole, the literature yields inconsistent results as to the program's success after kindergarten. For example, a study following Head Start children who took part in a Post-Head Start Transition program through the third grade found no achievement gains for the participants (Bickel & Spatig, 1999). However, it is reasonable to question whether the transition program elements were sufficient to maintain Head Start gains. Further, there is no control group with which to compare the progress of the children in the transition program to the progress of children not in the program or to those who had never attended Head Start. Finally, the limited sample in the Post-Head Start Transition program makes generalizations of any findings to a national sample of Head Start difficult.

Other data suggest that Head Start programs can have lasting effects. For example, one study found that female Head Start participants narrowed the gender gap in math (Kreisman, 2003). Findings also suggest that Head Start participation improves school readiness which can lead to enhanced school performance throughout elementary school (Abbott-Shim, Lambert, & McCarty, 2003, Lee, Brooks-Gunn, Schnur & Liaw, 1990). Data relating child care quality to positive child outcomes in third grade also point to potential benefits of Head Start participation (Burchinal, Roberts, Zeisel, Hennon, and Hooper, 2006; NICHD ECCRN, 2005). Considering that a national study of Head Start centers found that, on average, Head Start quality is on par with or better than alternate center-based child care options (Zill et al., 2003) it is reasonable to anticipate positive outcomes as a result of enrollment in the program.

Ongoing Longitudinal Efforts. Currently there are a number of longitudinal study efforts collecting valuable data regarding preschool and school-age experiences. Although not all of these studies focus exclusively on Head Start populations, all are collecting data that will allow examinations of the relationship between Head Start experiences and school-age outcomes. The studies are the Head Start Family and Child Experiences Survey (FACES), the Early Head Start Research and Evaluation Project (EHS), the Early Childhood Longitudinal Survey – Kindergarten Cohort (ECLS-K), and the Early Childhood Longitudinal Survey – Birth Cohort (ECLS-B). While these studies have provided and will continue to provide valuable data regarding the variety and nature of experiences of young children and their families as well as the relationships between those early experiences and later outcomes, the HSIS and the Third Grade Follow-Up Study will expand upon the knowledge gained from these combined research efforts.

The HSIS and the Third Grade Follow-Up Study provide the opportunity to explore questions related to Head Start using a nationally representative sample. The relevance of findings reported by these studies will not be biased by programmatic anomalies or limited by regional sample characteristics but will be applicable to the whole of the Head Start population. Further, the experimental design of the study allows for the comparison of children and families whose only significant difference is the treatment in question (i.e. access to Head Start). The preliminary results from the first year report show that Head Start increases 3-year-old children’s cognitive and social emotional development and children’s health as well as positive parenting practices (all the domains examined in the study). Impacts were found on some measures in each of these four domains. Findings were also positive, though less prevalent, for 4-year-olds. The lasting effect of these impacts has not yet been examined at the kindergarten or first grade level. The Third Grade Follow-Up Study will provide the opportunity to assess whether these effects were maintained or diminished or whether new effects occur.

Third Grade Environment. The onset of third grade highlights big changes in children’s understanding of their world as well as changes in the classroom instruction that they will receive. The children assessed as third graders will see the world in a markedly different way than they did just two years prior. Cognitively, most students at this age move from Piaget’s preoperational to concrete operational stage. Hallmarks of the concrete operational stage include more logical thought processes and the ability to take others’ perspectives into account (Piaget, 1983).

Evidence from the NICHD Study of Early Child Care and Youth Development leads us to expect a great deal of variability between the third grade classrooms of sample children.

However, on average instruction will focus on basic skills with special emphasis given to literacy skills. Furthermore, though structural classroom factors were somewhat stable from first to third grade in the NICHD study, aspects such as teacher sensitivity and number of literacy and math activities were not strongly related across that span (NICHD Early Child Care Research Network, 2005).

No Child Left Behind. Further, the Third Grade Follow-Up Study will be conducted in the context of No Child Left Behind (NCLB) implementation. The passage of the NCLB Act in 2001 attached consequences to its push for standards and accountability in public schools. Especially vulnerable to the potential consequences of NCLB are schools receiving Title 1 funding. As of 2002, 58% of all public schools, including 67% of elementary schools, received Title I funds (U.S. Department of Education, 2002). It is likely that many children from the Head Start sample will find themselves in such schools as Title I funds are targeted at low-income children.

Against the backdrop of NCLB requirements, many schools are experiencing powerful pressure to improve test scores. By the time the third graders in the Third Grade Follow-Up Study are assessed, schools will be required to test students annually in reading and language arts, and mathematics. Schools that have been designated as “in need of improvement” are likely to be experiencing particular pressure from administrators and parents to produce good test scores. Thus the Third Grade Follow-Up Study will have to take into consideration the possible effects on the findings of this test focus. For example, some schools may implement extra services or supports for third graders to help improve their test scores while others may not. In addition, data collection might be complicated by schools and teachers resistant to study demands on students’ time, even if few students are being assessed as part of the study. Assessing the impact of access to Head Start on children’s school performance will require consideration of their varied school experiences and the initiatives that individual schools are implementing. At the same time, analytic approaches incorporating school experiences may require foregoing true “experimental” analyses (i.e., those that compare program group children to control group children with no subgroups based on experience occurring post-random assignment).

Purpose of the Study

The primary purpose of the Third Grade Follow-Up Study is to answer questions about the longer term impact of Head Start through children's third grade year, including for whom and under what circumstances these impacts differ. This study will provide the opportunity to assess whether early differences between the treatment and control children were sustained through the first 4 years of school. The first-year results of the HSIS show effects on skills such as letter identification and spelling but not on oral comprehension, with effect sizes ranging from approximately 0.1 to 0.3. Some contend that for there to be a lasting effect of preschool outcomes, larger effects and higher order skills need to be achieved. Controversy about fade out of impacts goes as far back as Head Start's beginnings (M. Wolff and A. Stein, 1996). By measuring treatment and control cognitive differences, this study will help inform the fade out question.

In assessing third grade impacts, it is necessary to consider what new constructs and new skills should be assessed in the children and families in the study. There is concern that the biggest learning problem after the third grade is that disadvantaged students do not understand how to deal with ideas, generalizations, or abstractions as a result of a lack of experience in talking with adults about ideas (Pogrow and Stanley, 2000). Pogrow described this as students hitting a cognitive wall as they proceed in school and the curriculum continues to become more complex. Understanding whether the cognitive gains associated with access to Head Start were sufficient to impact children's comprehension skills is a key issue for this study.

It will also be important to identify and measure the dominant factors affecting third grade performance. These factors include child characteristics, parent characteristics and practices, home environment, as well as an accumulation of pre-school and school experiences. A particular challenge of all evaluations is to sort out the role each of these variables plays in mediating children's outcomes. While the experimental design of this study helps to clarify causal paths from the intervention to the outcomes, and with respect to factors established prior to random assignment such as child and family demographic characteristics, many of the variables of interest concern factors that emerged after random assignment (e.g., school experiences and child-parent interactions). This complicates the ability to isolate cause and effect relationships across the range of important influences on child development, Head Start among them. However, data acquired during this study may illuminate which factors contribute to the sustained impact of access to Head Start and which factors detract from it. It is likely to be the interaction of multiple variables that are most predictive of outcomes rather than simple linear relationships.

Research Questions

The Advisory Committee on Head Start Research and Evaluation (1999) recommended that the Head Start Impact Study address the following research questions: 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?

In the development of the HSIS, the research questions were restated into a set of overall program impact questions, and a set of questions that focus on the relationship between program impacts and children's experiences after random assignment (Westat, 2005). The preliminary research questions for the Third Grade Follow-Up Study follow this format, with some adaptations to reflect the focus on third grade outcomes.

- **Direct Impact of Access to Head Start on Children's Outcomes.** These questions address whether children's third grade outcomes differ depending on whether a child had access to Head Start. What is the impact of prior access to Head Start on children's cognitive development at the end of third grade? What is the impact of access to Head Start on children's social-emotional development at the end of third grade? What is the impact of access to Head Start on children's health status at the end of third grade? Does this vary by subgroups?
- **Potential Indirect Impact of Head Start on Children Through Direct Impact on Parents.** What is the impact of prior access to Head Start on parents' practices and support of their children's education by the end of third grade? For example, do the parents of third grade children who had access to Head Start report different parenting practices than the parents of third grade children who did not have access to Head Start? Do parental practices in support of their children's education (e.g., reading to child, taking an active role in their education) vary by subgroups?
- **Impacts on Experiences and Services.** What is the impact of prior access to Head Start on children's educational experiences and comprehensive services during the early school years? For example, are the characteristics of the schools that are attended by third grade children who had access to Head Start different from schools attended by children who did not have access to Head Start? Does this vary by subgroups?
- **Linking Experiences to Impacts and Outcomes.** How do the estimated impacts of access to Head Start vary by the nature and type of children's Head

Start experiences? If access to Head Start has an impact on the nature and type of children's experiences, does this impact lead to an impact on child outcomes? In addition, how are third grade outcomes influenced by parenting practices or school characteristics? What are the pathways through which Head Start access influences third grade outcomes? For example, to what extent does the impact of Head Start (e.g., months of attendance) depend upon children's early experience or to third grade experiences? These questions explore causal linkages between children's experiences after random assignment and children's later outcomes, to determine whether any impact on third grade outcomes is because of Head Start's impact on children's early school experiences (in the areas of parenting and school characteristics).

The Third Grade Follow-Up Study is an extension of the work conducted for the Head Start Impact Study that was based on the Congressional mandate that DHHS conduct research to determine, on a national basis, the impact of Head Start on the children it serves. This study will build on the HSIS randomized design to examine the longer term impact of Head Start at the third grade level.

A.2 How the Information Will Be Collected, By Whom It Will Be Collected, and For What Purpose

The original HSIS design called for collecting comparable data on two cohorts of newly entering children (a three-year old cohort and a four-year old cohort) and their families who were randomly assigned to either a treatment group (enrolled in Head Start) or a control group (that were not enrolled in Head Start, but were permitted to enroll in other available services in their community selected by their parents or be cared for at home). To draw the national sample, all eligible grantees/Delegate agencies were clustered geographically with a minimum number of eight grantees/delegate agencies within each cluster. The clusters were grouped into 25 strata based on state pre-K and childcare policy¹, race/ethnicity of the Head Start children served, urban/rural status, and region. Next, one cluster with probability proportional to the total enrollment of three- and four-year olds in the cluster, was selected from each stratum and approximately three grantee/delegate agencies were randomly selected from each cluster. From each of the 75 randomly selected grantees/delegate agencies, approximately 48 children per grantee/delegate agency were assigned to the Head Start treatment group and about 32 children were assigned to the control group. Sample children could not have been previously enrolled in Head Start. To avoid a sample size shortfall, small centers on the frame were grouped together within a program to perform center groups, each center group with a combined reported first year

¹ Data to define these strata were obtained from Children's Defense Fund (1999), *Seeds of Success* report.

enrollment of at least 27 children. The selection and random assignment of approximately 5000 children occurred during the Spring/Summer of 2002. The distribution of children by cohort or age group and by status (treatment or control group) are presented in Table 1.

Table 1. Number of Children in the Head Start and Non-Head Start Groups by Age Cohort

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

The Third Grade Follow-Up Study will follow the HSIS sample into the third grade. Data collection for this study will occur in the year in which most children in a cohort are in third grade (i.e., for the 4-year-old cohort, in fall 2006 to spring 2007, and for the 3-year-old cohort, in fall 2007 to spring 2008). Many challenges are presented by a longitudinal study with a national sample. Over time, families move and become more difficult to locate. Others, while they can be tracked successfully, move to distant locations that make in-person interviews and assessments difficult and costly. Respondent fatigue may occur where families begin to tire of the study after having been active participants for several rounds of data collection. Interviewers may feel that the study offers them few new challenges as they complete the same data collection tasks each year. New strategies are needed to keep the study fresh and exciting for respondents and interviewing staff. We have been successful in gaining high cooperation from respondents for five rounds of HSIS data collection and four rounds of tracking. High response rates were achieved for the parent interviews and the child assessments as presented in Table 2.

Table 2. Percent of HSIS Parent Interviews and Child Assessments Complete by Data Collection Period

Percent Complete									
			Tracking		Tracking		Tracking		Tracking
	Fall 02	Spring 03	Fall 03	Spring 04	Fall 04	Spring 05	Fall 05	Spring 06 3-year old cohort	Spring 06 4-year-old cohort
Parent	86%	83%	84%	81%	83%	81%	83%	80%	82%
Child	80%	84%		81%		78%		77%	

N=4,667

Built into each wave of data collection for the Third Grade Follow-Up Study is an assumed attrition rate of 2-3 percent. This attrition rate is based on the HSIS experience and

includes refusals, children/families who move from the area, and children and families who could not be located.

The data collection plan for the Third Grade Follow-Up Study will build upon the successful procedures implemented in the original HSIS. Data collection will include waves in spring of the third grade years for both cohorts as well as tracking update information in spring 2007 and fall 2007 for the 3-year old cohort and in fall 2007 and spring 2008 for the 4-year old cohort. Child assessments and parent interviews will be conducted in the child’s home in the spring of the child’s third grade year in school. During the spring data collection, information will be collected from each study child’s school principal and teacher(s). School data, such as average class size, school improvement status, and state test scores will also be collected from secondary sources (e.g., CCD (Common Core of Data), etc.). Table 3 provides a summary of the data collection activities.

Table 3. Data Collection Schedule

School Year Cohort 1 (C-1) Cohort 2 (C-2)		2006-2007 Cohort 1 (second grade) Cohort 2 (third grade)		2007-2008 Cohort 1 (third grade) Cohort 2 (fourth grade)	
Data Source	Cohort	Fall	Spring	Fall	Spring
Children	C-1				X
	C-2		X		
Parent/Primary Caregiver	C-1		X*	X*	X
	C-2		X	X*	X*
Teacher	C-1				X
	C-2		X		
Principal	C-1				X
	C-2		X		

*denotes the parent update.

The Third Grade Follow-Up Study primarily is using instruments previously approved by OMB and used in the Head Start Impact Study (OMB #0970-0029), Family and Child Experiences Study (OMB #0970-0151), or the Early Childhood Longitudinal Study-Kindergarten. The proposed instruments are copyrighted, so only instrument descriptions are provided in Appendix A.

The third-grade child assessment instrument must build on the structure established for the HSIS data collection, accurately measure children's cognitive development throughout the span of the study, and be adaptive to the different developmental levels of third-grade children. Substantial changes in the instruments will limit the opportunity to do longitudinal analysis of impacts on growth. However, the existing HSIS instruments need to be adapted to focus on the skills and activities appropriate for third grade students. The instruments will continue to measure outcomes that are important from an educational point of view and are likely to be affected by Head Start exposure. Any new tests will have previously demonstrated reliability and validity for Head Start or other low-income populations.

The child assessment will focus on the areas of reading (language and literacy), mathematics, and executive functioning. In addition, children will be asked to respond to a self-report instrument that includes items about school, their attitudes, motivation, relationships and behavior. The total battery will take about 1 hour to administer. We will assess each child using the following instruments: Early Childhood Longitudinal Study-Kindergarten Cohort (ECLS-K) reading assessment, the Letter-Word Identification, Applied Problems, and Calculation subtests from the Woodcock-Johnson III Tests of Achievement, the Numbers Reversed and Auditory Working Memory from the Woodcock-Johnson III Tests of Cognitive Abilities, and the ECLS-K Self-Description Questionnaire (SDQ). In order to continue to measure Spanish language skills, children who were classified as bilingual in the HSIS will also be administered the Bateria Woodcock-Muñoz Identificación de letras y palabras. Spanish speaking children (Puerto Rico) will be administered the Identificación de letras y palabras, Comprensión de textos, Problemas aplicados, and Cálculo tests from the Bateria Woodcock-Munoz Tests of Achievement, the Inversión de números test from the Bateria Woodcock-Muñoz Tests of Cognitive Abilities, and translated versions of the Auditory Working Memory test from the Woodcock-Johnson III Tests of Cognitive Abilities and ECLS-K Self-Description Questionnaire.

The ECLS-K third-grade reading assessment built on the 1992 and 1994 National Assessment of Educational Progress (NAEP) Fourth-Grade Reading Frameworks and more difficult skills from the ECLS-K kindergarten and first-grade reading assessments. The ECLS-K third-grade assessment measures phonemic awareness, single word decoding, vocabulary, and passage comprehension.

The Letter-Word Identification test from the Woodcock-Johnson III Tests of Achievement measures letter and word identification skills. The easier items require the respondent to identify letters. As the items increase in difficulty, the respondent is asked to pronounce words correctly but the respondent is not required to know the meaning. Passage Comprehension requires the respondent to read a short passage and identify a missing key word that makes sense in the context of the passage. Applied Problems requires the respondent to analyze and solve math problems. The respondent is required to listen to the problem, recognize the procedure to be followed, and then perform simple calculations. Calculation measures the respondent's ability to perform mathematical computations. The Numbers Reversed test measures short term memory as well as working memory or attentional capacity. The Auditory Working Memory test measures auditory memory span as well as working memory or divided attention. These tests are included in the Woodcock-Johnson III Tests of Cognitive Abilities.

The Self-Description Questionnaire (SDQ) used with the third-grade ECLS-K cohort provides a self-report measure of the child's cognitive and socio-emotional attitudes. In this instrument, the child is asked to respond to a series of items such as "I enjoy doing work in all school subjects" or "I often argue with other kids" using a four-point response format (i.e., "not at all true, a little bit true, mostly true, and very true"). The 42-item SDQ includes both indirect cognitive measures (self-ratings of competence in reading [eight items], mathematics [eight items], and all school subjects [six items]) and socio-emotional questions related to peer relationships (six items) and problem behaviors (i.e., anger and distractibility (seven items) and sad, lonely, or anxious (seven items)).

The Bateria Woodcock-Muñoz tests administered to the Spanish and bilingual children are similar to the Woodcock-Johnson III tests administered to the English-speaking children.

The child assessment will be administered in the child's home by a Westat trained field assessor. The assessor will enter the child's responses into a laptop computer using the Computer-Assisted Personal Interview (CAPI) technology. The SDQ will be designed as an Audio Computer-Assisted Self Interview (ACASI). These technologies are discussed in the next section. The SDQ will be available in Spanish using the ACASI technology.

The parent interview builds upon the HSIS parent interview that includes both parent report of child outcomes and parent outcomes and will take about one hour to administer. The parent interview for the Third Grade Follow-Up Study will include the following cognitive and socio-emotional child outcomes: Parent-Reported Emergent Literacy Scale (PELS); Social Skills and Positive Approaches to Learning (SSPAL); Social Competencies Checklist (SCCL); Total Problem Behavior (TBP); and Parent-Child Relationship. The parent interview also includes health and parenting outcomes such as the following: whether the child has health insurance, whether the child has received dental care, child's health status, whether the child needs ongoing medical care, whether child received medical care for an injury in the last month, educational activities with child, amount of reading to the child at home, cultural enrichment activities with child, use of physical discipline, school communication, parenting styles, summer learning activities, and parent involvement. Demographic information as well as information on income, housing, and neighborhood characteristics are also collected during the parent interview.

The parent interview will be conducted in the parent's/primary caregiver's home by a Westat trained field interviewer who will enter the parent's'/primary caregiver's responses into a laptop computer using CAPI technology. The parent interview will be available in both English and Spanish. For parents/primary caregivers who speak neither English nor Spanish, an interpreter will be used.

Ongoing tracking in longitudinal studies is critical to maintaining high response rates. The tracking update form will be used to verify and update if necessary school and contact information. Tracking updates will occur in spring 2007 for the 3-year old cohort, in spring 2008 for the 4-year old cohort, and in fall 2007 for both cohorts. The tracking updates will primarily be conducted over the telephone with in-person follow-up as necessary. Tracking updates will take about 10 minutes to complete.

We will gather specific information about the child's experiences and development from the perspective of the teacher. The self-administered teacher survey will include questions to obtain biographical information including education and years of experience, inquiries regarding program elements, quality of management, belief scales to assess staff attitudes on working with and teaching children, and the operation and quality of the program. Items on literacy promoting activities, parallel to questions used in the ECLS-K, are included in the teacher instrument. Use of these items provides a national sample benchmark for the measures. The teachers will also be asked to rate each study child in their classroom using the self-administered teacher/child report form. Information will be collected in the following areas: teacher-child

relationship, classroom activities, general background, academic skills, school accomplishments, and health and developmental conditions or concerns. A teacher/child rating form will be completed for each study child's math and reading/language arts class. For additional study children in a class, the teacher will need to complete a separate teacher/child report form for each child. The self-administered teacher survey and teacher/child report form for one child will require about 30 minutes to complete.

The school principal is another source of data for school demographic characteristics and quality indicators for the school, teachers, and classrooms. The self-administered principal survey will take about 20 minutes to complete and will provide information on the overall operation and quality of the program, including teacher performance, staffing and recruitment, overall staff qualifications, teacher education initiatives and staff training, parent involvement, curriculum and assessment, and demographic information.

We will rely on information from secondary sources, such as the Common Core of Data (CCD) or the school or district website, to track a school's record with respect to such issues as attendance, disciplinary issues, immunizations of children, average test scores, number of children receiving free or reduced school lunch, school improvement status, and teacher/student ratio.

The information collected for the HSIS will be used to measure the impact of Head Start through first grade on the children it served. The Third Grade Follow-Up Study will provide a measure of the impact of Head Start on children through the third grade.

A.3 Use of Automated, Electronic, Mechanical, or Other Technological Collection Techniques

Westat plans to use a CAPI (Computer-Assisted Personal Interview) instrument for the child assessment with a separate ACASI (Audio Computer-Assisted Self Interview) component for the Self-Description Questionnaire. While we used a paper instrument for the original HSIS, we believe that switching to a computer-based approach for the Third Grade Follow-Up Study has certain benefits. The child assessment for the study will include Woodcock-Johnson III subtests that require the establishment of basals and ceilings. A child starts the test at a particular point depending on the child's grade. The basal criterion is met when the child

answers the six lowest-numbered items correctly. If the basal is not obtained in that section, the interviewer/assessor must test backward until the child meets the basal criterion. Once the child meets the basal criterion, the interviewer/assessor then skips to the point at which the test was interrupted. The interviewer/assessor continues to administer the test until the child gives six consecutive incorrect answers, thereby establishing the ceiling. This approach limits the number of questions that need to be asked but can be tricky for the interviewer/assessor to administer correctly. We believe errors in skipping to the wrong section will be avoided by computer-based scoring which directs the interviewer/assessor to the correct section.

The Self-Description Questionnaire is designed to be self-administered to children. We will program this instrument as an ACASI so that the children have the questions read to them by the computer rather than having to read the questions by themselves. We will first use a short tutorial to teach the child how to use the computer while the interviewer/assessor observes to make sure that child can master the task. Following this tutorial, the interviewer/assessor gives the child headphones and moves across from the child to allow the child privacy to complete the self-administered questionnaire.

We believe that using an ACASI with third grade children has several advantages over a paper and pencil interview that is either self-administered or interview-administered. First, using the ACASI addresses concerns about the varied literacy of the children. Because the computer reads the questions and answers (reading each answer as it is highlighted on the screen), children who are poor readers or nonreaders can still complete the instrument.

A second reason to prefer the ACASI is that it gives the child greater privacy to answer the questions and may lead to more honest answers. Some of the questions on the Self-Description Questionnaire are sensitive in nature and previous research has demonstrated the utility of CASI in soliciting more honest answers to sensitive behaviors (Currivan, Nyman, Turner, and Biener, 2004; Des Jarlais et al., 1999; Metzger et al., 2000; Turner, Ku, Rogers, Lindberg, Pleck, and Sonenstein, 1998).

A third reason to introduce the ACASI at this point is because we believe administration using the computer is more likely to capture the children's interest than a paper assessment. As third grade students, most children will be experienced with computers from school and/or home. We anticipate that they will be able to use the computer with few problems, as has been our experience on other studies with similar populations. Using a computer is an age-appropriate method. We have also found that the ACASI is a faster method of administration than

an interviewer-administered method. As the study children have been completing the child assessments now for several years, a fresh approach is required to keep the assessments from becoming stale. An additional advantage to using ACASI is that all children hear the exact same voice reading the questions and answers, which provides for greater standardization.

Because we will use a computer to administer the child assessments, we also will program the parent interview as a CAPI so that the interviewer has just one mode to use for both instruments. Programming the parent instrument as a CAPI ensures that skips are correctly administered and all questions are answered. In addition, it eliminates the need for editing by the site coordinator and coding and data entry at the home office. The CAPI will be programmed in both English and Spanish.

Using ACASI for the Self-Description Questionnaire and CAPI for the parent interview and the child assessment cognitive subtests will reduce the time the interviewers take to verify that each instrument was administered correctly and thus reduce respondent burden.

A.4 Efforts to Identify Duplication

In the late 1990's, the US General Accounting Office (GAO) released two reports concluding that (1) "...the Federal government's significant financial investment in the Head Start program, including plans to increase the number of children served and enhance the quality of the program, warrants definitive research studies, even though they can be costly" (U.S. General Accounting Office, 1998) and (2) this information need could not be met because "...the body of research on current Head Start is insufficient to draw conclusions about the impact of the national program" (U.S. General Accounting Office, 1997).

One purpose of the Head Start Impact Study was to measure the impact of Head Start on children's early development and school readiness. The Third Grade Follow-Up Study allows us to build upon the existing HSIS design and determine the longer term impact of Head Start on the well-being of children and families. Specifically, the study will examine the degree to which the impacts of Head Start on initial school readiness are maintained or changed by children's school experiences and the various school quality and family/community factors that come into play up to and during third grade. Other studies have examined third-grade performance and growth but no study has measured the impacts of the Head Start program at the third grade level using a randomized design and a nationally representative sample.

A.5 Minimizing Impact on Small Businesses or Other Small Entities

No small businesses or other small entities will be involved in the data collection for the Third Grade Follow-Up Study.

A.6 Consequences If the Collection Is Not Conducted

As recommended by the Government Accounting Office and mandated by Congress, “definitive research studies” are legislatively required to assess the effectiveness of Head Start nationally on the school readiness of participating children. Despite increasing expenditures, including an appropriation of \$6.84 billion in fiscal year 2005, “the body of research on current Head Start is insufficient to draw conclusions about the impact of the national program.” The Head Start Impact Study provided the data to allow such an evaluation through first grade. The Third Grade Follow-Up Study will provide data for determining longer term impacts. If this study is not conducted, there is not a current mechanism determining the long term impact that Head Start has on enhancing a child’s school readiness and performance.

A.7 Special Circumstances

This Third Grade Follow-Up Study will be conducted in a manner entirely consistent with the guidelines in Title 5, Section 1320.6 of the Code of Federal Regulations. There are no special circumstances that might require deviation from these guidelines.

A.8 Consultation with Persons Outside the Agency

The public announcement for the Third Grade Follow-Up to the Head Start Impact Study was published in the **Federal Register** on Tuesday, August 15, 2006 (Vol. 71, No. 157, pp 46916-46917). The text of the announcement is contained in Appendix B. A single comment was received from the public through this notice. That comment, and our response, is included in Appendix C.

Information concerning the Head Start Impact Study was included in the “Report of the Advisory Committee on Head Start Research and Evaluation” as part of the Head Start

Amendments of 1998. The Head Start Act was reauthorized, through fiscal year 2003, by the Coats Human Services Amendments of 1998, PL 105-285 (10/27/98). Consultation meetings were held with Advisory Committee on Head Start Research and Evaluation on January 12, 2001, June 16-17, 2003, and September 28-29, 2005. (Advisory Committee Members are listed in Appendix D.) In addition, a consultation meeting was held with the Advisory Committee on Head Start Accountability and Educational Performance Measures on June 16, 2005. The purpose of these meetings included providing and discussing information about the proposed design and its implementation, instrumentation, and analysis and reporting, as well as general advice from the Advisory Committee Members.

In addition to the Advisory Committees, Westat assembled a consultant cadre to assist the project team in the development of the initial design for the Head Start Impact Study in the areas of assessment of Spanish-speaking children, socio-emotional development, language and literacy, parenting skills and activities, comprehensive services, educational components and observations, and statistics and analysis. Consultants were used to augment the skills and experiences of the project team on particularly difficult technical and substantive issues. Most consultation meetings were conducted as conference calls, however, an analysis consultation meeting was convened on May 7, 2003 to assist the project in designing the complicated analysis plans.

A list of suggested consultants for the Third Grade Follow-Up to the Head Start Impact Study have been submitted to the Program Officer. To date, Westat staff have met with the following consultants to discuss the overall research and design issues for this study:

Mark Greenburg, Pennsylvania State University
Robert Pianta, University of Virginia
Marty Zaslow, Child Trends

A.9 Remuneration to Respondents

In order to minimize the burden placed on families for participating in the study and to maximize response rates, we will provide parents with \$50, provided in two installments, for completing both a tracking interview and complete parent interview. A gift card or cash will be provided to the parent upon completion of their interview (\$30) and each tracking update (\$20). Tracking updates are only conducted during the time periods when full interviews are not

administered. A family can receive no more than \$50 in two installments over one school year for completing both a tracking interview and complete parent interview. This is an increase in the remuneration for the parent interview but it is important to underscore the fact that this is a new and different study from the original HSIS for families who had begun to experience respondent fatigue. The new incentive should emphasize the importance of the new study and the value of the respondent's participation in it. We will use the same remuneration plan for the teachers as was implemented in the HSIS (i.e., completed teacher survey and teacher/child rating form(s)). The remuneration was graduated depending on the number of child forms completed: \$15 for the teacher survey and 1-3 child forms, \$25 for the survey and 4-10 child forms, and \$35 for the survey and 11+ child forms. We anticipate that most teachers will receive an incentive of \$15. We do not plan to offer remuneration for the principal survey as we do not feel that an incentive is necessary for administrators at this level. Children will receive a non-cash incentive not to exceed \$5 to enhance their interest, increase motivation, and ensure high rates of participation. Table 4 provides a summary of the data collection instruments, activities, and incentives.

Table 4. Data Collection Instruments and Activities

Instrument	Respondent or Setting	Activity	Cash/Gift Card Incentive	Conducted by	Time Required
Parent Interview	Parent/Primary Caregiver	In-person Interview	\$30	Interviewer/ Assessor	1 hour
		Tracking Update Telephone Interview	\$20		10 minutes
Child Assessment	Child	Individual Assessment	Non-cash incentive not to exceed \$5	Interviewer/ Assessor	1 hour
Teacher Survey and Teacher Child Rating Reports (TCRs)	Teacher (Third Grade Reading and Math Teacher)	Self-administered Survey Individual Child Ratings 1 - 3 children 4 - 10 children 11 + children	\$15 \$25 \$35	Delivered and Collected by Interviewer/ Assessor	30 minutes for the teacher survey and the teacher/child rating form
Principal Survey	School Principal	Self-administered Survey	\$0 (no remuneration)	Delivered and Collected by Interviewer/ Assessor	20 minutes

A.10 Assurances of Privacy

All Westat staff members sign the Westat pledge of privacy for the study. In addition, all field staff signed a privacy pledge.

For some parent respondents, the issue of privacy of information, particularly relating to address and telephone information collected for later tracing of respondents, is a matter of great concern. Participants will be assured that the information collected will be used for research purposes only by the research team, and that contact name and address information and other survey data will not be given to bill collectors, legal officials, other family members, etc. Also, names will not be linked to the survey in any way. Similarly, program staff will be assured that no information on individual schools or classrooms, including the identity of individual teachers or principals, will be released.

We will implement procedural steps, similar to the steps used in the HSIS, to increase respondent confidence in our privacy procedures. We will generate a set of identification labels with a unique respondent ID number and bar code. These labels will be affixed to each of the data collection instruments for a respondent. The use of bar codes in conjunction with the numbered identification labels enables the receipt control staff to enter cases by reading the bar code with a wand, making receipt of completed interview packages also more efficient.

A.11 Questions of a Sensitive Nature

In order to achieve the goal of enhancing the cognitive and social competencies of children from low-income families, Head Start needs to understand the social context within which Head Start children and their families live as the child progresses through school and the nature of the daily challenges that they face. Thus, several questions of a sensitive nature are included in the parent interview.

Queries of a sensitive nature include questions about feelings of depression, use of services for emotional or mental health problems or use of services for personal problems such as family violence or substance abuse. The questions obtain important information for understanding family needs and for describing Head Start's long term impact in these aspects of individual and family functioning.

Questions about the respondent's neighborhood may also be sensitive but are important to obtain information about the contextual factors in communities that impede or facilitate family well-being.

Another set of questions of a sensitive nature focuses on the families' involvement with the criminal justice system and the child's exposure to neighborhood or domestic violence. Although highly sensitive, this information is crucial to understanding family needs, identifying risk factors for the child's development and fully describing the contextual factors in families that impede or facilitate family well-being. This issue was given a high priority by the National Academy of Sciences (*Beyond the Blueprint: Directions for Head Start Research*) in formulating recommendations for Head Start research initiatives. A full understanding of these issues is essential both for Head Start program planning and for a realistic assessment of what Head Start can achieve over the long term.

The purpose of the interview and how the data will be used will be explained to all participants. Participants will be reassured in person and in writing that their participation in the study is completely voluntary. A decision not to participate will not affect their standing in any government program, and if they choose to participate, they may refuse to answer any question they find intrusive. All individuals' interview responses will be private and none of their answers will be reported to any program, agency, or school but will be combined with the responses of others so that individuals cannot be identified. All interviews will take place in a setting where the respondent's privacy can be assured.

The voluntary nature of the questions and the privacy of the respondent's answers will be restated prior to asking sensitive questions. In all cases, questions on these topics are part of a standardized measure or have been carefully pretested or used extensively in prior studies with no evidence of harm.

A.12 Respondent Burden

Tables 5a-c present data on the total burden for respondents to the Head Start Impact Study for each data collection point. For the two years of data collection the average number of responses is 12,850 and the average annual hours requested is 6,873 (as shown in table 5c).

Table 5a. Estimated Response Burden for Respondents in the Third Grade Follow-Up to the Head Start Impact Study-Spring 2007 (Year 1)

INSTRUMENTS	NUMBER OF RESPONDENTS	NUMBER OF RESPONSES PER RESPONDENT	AVERAGE BURDEN HOURS PER RESPONSE	TOTAL BURDEN HOURS
Parent Tracking Interview	4,667	1	1/6	778
Parent Interview	1,720	1	1	1,720
Child Assessment	1,720	1	1	1,720
Teacher Survey/TCR	2,580	1	1/2	1,290
Principal Survey	1,462	1	1/3	487
Totals for Spring 2007	12,149			5,995

Total Respondents for Year 1: 12,149

Total Responses for Year 1: 12,149

Total Burden Hours for Year 5,995

Table 5b. Estimated Response Burden for Respondents in the Third Grade Follow-Up to the Head Start Impact Study-Fall and Spring 2008 (Year 2)

INSTRUMENTS	NUMBER OF RESPONDENTS	NUMBER OF RESPONSES PER RESPONDENT	AVERAGE BURDEN HOURS PER RESPONSE	TOTAL BURDEN HOURS
Parent Tracking Interview	4,667	1	1/3	1556
Parent Interview	2,042	1	1	2,042
Child Assessment	2,042	1	1	2,042
Teacher Survey/TCR	3,063	1	1/2	1,532
Principal Survey	1,736	1	1/3	579
Totals for Spring 2008	13,550			7,751

Total Respondents for Year 2: 13,550

Total Responses for Year 2: 13,550

Total Burden Hours for Year 2: 7,751

Respondents

Table 5c. Average Annual Estimated Response Burden for Respondents in the Third Grade Follow-Up to the Head Start Impact Study-Fall and Spring 2008 (average across both years of the study)

INSTRUMENTS	NUMBER OF RESPONDENTS	NUMBER OF RESPONSES PER RESPONDENT	AVERAGE BURDEN HOURS PER RESPONSE	TOTAL BURDEN HOURS
Parent Tracking Interview	4,667	1	1/4	1167
Parent Interview	1881	1	1	1881
Child Assessment	1881	1	1	1881
Teacher Survey/TCR	2822	1	1	1411
Principal Survey	1599	1	1/3	533
Average Annual Totals	12,850			6873

Total Average Respondents across Both Years: 12,850

Total Average Responses across Both Years: 12,850
Total Average Burden Hours across Both Years: 6,873

	Annual Avg
Child Assessment	1
Principal Survey	1
Parent Assessment	1
Parent Tracking	4
Teacher Survey	2
	12

SUPPORTING STATEMENT

Year 1	12
Year 2	13
Average	128

A.13 Total Annual Cost Burden

There are no direct monetary costs to participants other than their time to participate in the study.

A.14 Annualized Cost to the Government

The total cost to the Federal Government for the Third Grade Follow-Up to the Head Start Impact Study, under the terms of the contract to Westat is \$9,594,875. These costs include development of the project materials, the use of consultants, data collection, data coding and analysis and preparation of the necessary reports and presentations. Respondent expenses and incentives are included in these costs. The budget breakdown by year follows: Year 1 - \$3,663,719; Year 2 - \$4,374,500; and Year 3 - \$1,556,656.

A.15 Reasons for Any Program Changes

There are no planned program changes. OMB will be promptly notified if any changes are recommended by outside consultants or identified by project staff.

A.16 Plans for Tabulation and Statistical Analysis and Time Schedule

Analysis Plan

The analysis of the Third Grade Follow-Up Study data will begin with an analysis of children's educational experiences from the third grade through the initial random assignment at the time of application to Head Start. This will consist of both descriptive analyses and experimental estimates of the impact of access to Head Start on the nature and quality of children's early school experiences, should Head Start influence which schools and classrooms children attend during their early elementary school years. The descriptive piece will be enriched, to capitalize on the opportunity that will now emerge, and that in general is quite rare, to track a large sample of disadvantaged children through several years of early elementary school experience. These analyses will help the reader understand the context within which the many important influences on child development—including Head Start's impact—guide students' social and intellectual development during the critical kindergarten to third grade years.

We could also envision that Head Start would *change* the school experience of the children given access to the program in their pre-school years. The HSIS consisted of two groups of randomly assigned children—a 3-year and a 4-year old cohort of Head Start eligible children who had no prior Head Start experience at the time of random assignment. The 4-year old cohort had the opportunity for, and was observed, during one year of “preschool experience” followed by kindergarten, first and now third grade in elementary school; on the other hand, the 3-year old cohort had the opportunity for, and was observed, during 2 years of “preschool experience” followed by kindergarten, first and now third grade in elementary school. Random assignment at the time of application to Head Start to either the treatment or control group can have an impact on child and parent outcomes by changing **where** children go (i.e., what we call their “setting”) as well as **what** they experience once they are in a particular setting. For example, parents may have made different decisions about their child’s preschool experiences as a consequence of being assigned to the control group. These changes, or other shifts in parental perspectives, could also influence where to send a child **for elementary school**, meaning that access to Head Start could alter the identity of the schools and classrooms children attend once they enter kindergarten and hence the nature and quality of that school experience. The first focus of our analysis of Head Start’s impact involving the Third Grade Follow-Up Study data will explore whether treatment status influenced how school experiences evolved over time for both age cohorts through the end of third grade.

These analyses will consist of the preparation of tables such as the one illustrated (using hypothetical data) in Table 6. In this example, access to Head Start has an impact on the quality of children’s elementary school teacher, possibly by increasing the salience of education to primary caregivers and thus better preparing parents to advocate for a particular teacher for their child at the school s/he attends, and by possibly inducing some parents to select different schools for their children (e.g., non-sectarian or religiously affiliated private schools, charter schools, or even obtaining school vouchers where available).² Should such events take place as a consequence of Head Start, their cumulative effect on the quality of the elementary education received by former Head Start students would have obvious policy importance (for both Head Start and the public schools that serve Head Start participants) and could strongly affect the mediational influence of the school experiences on the long-run benefits to children of early participation in Head Start.

² Some questions have been raised as to whether such options are, in fact, available to low-income parents. As noted in a recent article by Teske, Wolf, and Hill (Teske, P., P. Wolf, and P. Hill, (2006), “Poor Parents can be Good Choosers.” *Education Week*, April 19, 2006.), “Like middle-class parents who have always had choices, low-income parents don’t look for alternatives if their children are happy and successful in school. But once they start thinking about school options, low-income families want information about schools and think hard about the choices they have. Poor parents seek to escape problems evident in their children’s current school, and have definite ideas about the differences between one child and another (our studious boy, our distractible girl) that lead them to search for an appropriate match between child and school.”

As shown in the table, access to Head Start (hypothetically) leads 12 percent more of the children to be in elementary school classrooms with high quality teachers, and 6 percent more to be in classrooms with medium quality teachers. Because all of the children from the program and control groups are included in this comparison, these shifts in exposure represent experimental estimates of the impact of Head Start on subsequent in-school experiences. Such analyses can be conducted on a wide range of school experience measures once all children

Table 6. Hypothetical Example: Impact of Head Start on Access to “Quality” Elementary School Teachers

Teacher Qualifications	Complete Program and Control Groups		
	Percent of Control (C) Group	Percent of Program (P) Group	Change Due to Access to Head Start (P-C Difference=Impact)
Have teacher with High qualifications	54%	66%	12%
Have teacher with Medium qualifications	15%	21%	6%
Have teacher with Low qualifications	31%	13%	-18%
Total	100%	100%	

are in school (i.e., have data on each measure) to measure directly ITT (impacts on the treated) impacts of Head Start on such important school and classroom characteristics as the child-teacher ratio, quality of the instruction children receive, the availability of extra-curricular activities, etc.

Annual Impacts on Child and Parent Outcomes

After examining the nature of children’s in-school experiences that will condition all other outcomes for the study, we will turn to the analysis of cognitive and behavioral development and health, and Head Start’s impact on those outcomes, through the end of third grade. This will provide a longer-term perspective on how Head Start influences child and parent outcomes, extending the kindergarten and first grade analyses of the original HSIS into the core elementary school years. These analyses will deal with what we have referred to as the “main” intent to treat (ITT) impact estimates³ focusing on topics

³ This discussion focuses on the primary experimental impact estimates called *ITT estimates*. These are based on the use of the entire sample of children who were randomly assigned at the start of the study, and as such provide information about the impact of access to Head Start. That is, the estimates include some children who are offered enrollment but who do not actually participate in Head Start or participate for very little time. A second set of impact estimates, referred to as *impact on the treated (IOT)*, will also be calculated and presented to adjust for the fact that

inspired by the research questions recommended for the original study by the National Advisory Panel: “*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?*” These overarching questions lead to the following research questions as concerns the third grade experience and outcomes of the original study:

- **Direct Impact of Head Start on Children**
 - What is the impact of access to Head Start during the pre-school years on the extent of children’s **cognitive development** at the end of third grade? If an impact is found at the end of third grade, it can be due to the carryover consequences of school readiness gains in place as of the spring prior to kindergarten on third grade. Additionally, the impact may be due to the fact that children are fundamentally changed by the Head Start experience in ways that lead them to react differently to subsequent educational experiences. The ITT analysis will measure the combined effect of these two channels through which impacts may appear. A potential additional analysis, if there is interest, would be to try and disentangle the influences from these two separate sets of situations.
 - What is the impact of prior access to Head Start during the pre-school years on the extent of children’s **social-emotional development** at the end of third grade? Similar teasing out of the origins and channels of cognitive impacts could also be considered in the case of social-emotional impacts.
 - What is the impact of prior access to Head Start during the pre-school years on children’s **health status** at the end of third grade?
- **Potential Indirect Impact of Head Start on Children Through Direct Impacts on Parents**
 - What is the impact of prior access to Head Start on **parent’s practices and support** of their child’s education at the end of third grade?

some children assigned to the treatment group do not, in fact, avail themselves of the opportunity (called no shows) and some children assigned to the control group manage to find their way into Head Start at centers not operated by the grantees in the study sample (called *cross-overs*). The challenge of creating reliable IOT measures from experimental data in the presence of no-shows and crossovers has been recognized in the evaluation literature for some time. The Westat team examined several options for conducting these analyses and recommended to OPRE an approach adopted for the original evaluation that entails the computation of lower and upper bounds on the IOT impact of Head Start through high- and low-end assumptions about the outcomes that would have been observed for crossover children had they *not* participated in the program. Further, it deals with no-shows by assuming they experienced no impact of Head Start since they never participated and rescaling the findings to reflect impacts on just participants. We recommend that this same procedure be used to put bounds around the IOT impacts of Head Start from the Third Grade Follow-up, restricting the analyses as before to outcomes with natural growth trajectories (the instances in which an upper bound can be calculated) such as children’s cognitive development.

The Final Report from the HSIS will have addressed similar questions through the end of first grade examining all of the outcomes collected through that point. The extension of the study through the end of third grade will augment these early findings by doing the following:

- Allowing the examination of how children’s experiences, and impacts on child and parent outcomes, might change as children continue their education through the end of third grade; such changes could occur for example because higher order skills are measured in third grade.
- Expanding the range of possible measures of children’s experiences to encompass third grade school experiences, and new parent and child outcomes.

The types of analyses that will be used to answer these newly framed questions collected are described below.

To ensure consistency with the original study, estimates of the main impact of access to Head Start will be estimated for all outcome measures collected at the end of third grade and will be derived using, the same basic statistical model used by the initial study to estimate impacts through the end of first grade:

$$(1) Y_i = \delta_0 + \delta_1 Z_i + \delta_2 W_i + \delta_3 A_i + e_i \text{ where}$$

- Y_i = spring outcome measure for individual i
- Z_i = binary treatment variable with $Z_i=0$ if the student is assigned to the control group and $Z_i = 1$ if the student is assigned to the treatment group,
- W_i = background characteristics of individual i used as covariates, from Fall 2002 data collection (e.g., race/ethnicity, gender)
- A_i = age of individual i on day of outcome data collection, in days
- e_i = random error term attributable to sampling variation.

The coefficient on the treatment variable, δ_1 , measures the effect of access to Head Start on the outcome variable Y for the average child in the treatment group—the ITT impact estimate. The background variables in W increase the statistical precision of the model and hence the impact estimate, while the inclusion of A_i “neutralizes” any developmental difference between treatment and control group children at the time Y_i was measured. The particular set of background characteristics used as covariates in the model, the W_i , will parallel those used in analyses conducted through the end of first grade but will be refined to any other background variables thought to be important conditioning factors for child and parent outcomes at an older age. All will still need to have been measured at the time of random

assignment in 2002, such as child gender, child race/ethnicity, whether mother was a teen when she gave birth to the study child, whether the child lives with both biological parents, mother's highest level of education, whether mother was a recent immigrant, home language, mother's marital status, child's age in months as of September 1 2002, parent/primary caregiver's age, time of fall 2002 testing, and the baseline measure of the outcome variable.

Where available, the fall 2002 baseline measure of the outcome variable Y will also be entered into the equation, since it too can explain a lot of the child-to-child variability in Y and hence increase the statistical precision of the impact estimate. The residualized version of the "pre-test" value of Y will be used for this purpose, to assure that no early effects of Head Start are removed from the longer-term estimates should late (i.e., post-random assignment) collection of this measure in Fall 2002 lead to small treatment-control differences in the measure due to very early program impacts.⁴ The residualizing assures that these impacts are not removed in calculating later program effects where the pretest is used as a covariate in the model.

For all continuous outcome measures (the majority of the measures) the model of impact in Equation (1) will be estimated through OLS regression using PROC REGRESS in SUDAAN (which allows the correct calculation of standard errors from a complex sampling design). For those measures that are dichotomous in nature (e.g., whether the student was retained in grade or not), logistic models will be estimated using PROC LOGIST, and for the few categorical outcomes (e.g., different "levels" of parental reading to their child), the models will be estimated using PROC MULTILOG which provides both an overall estimate of impact and impacts by the separate categories of the particular outcome variable.

In all cases, separate analyses will be conducted by age cohort. As discussed below, we will examine the extent to which impacts differ across age cohorts and for the combined sample as well as, separately, the Puerto Rico sample. Weighted data will be used for all analyses to accurately represent the national Head Start system.

⁴ The residualizing will be accomplished by (1) regressing each fall outcome measure on a constant and the treatment dummy variable, and then (2) producing a predicted value for the baseline measure that is then subtracted from the actual value of the measure to get a residual for each child.

Longitudinal Analysis for Impact Analysis

For those outcomes that have been collected at multiple time points, and that are assumed to have a natural growth pattern, repeated measures analysis will be used pooling the new third grade observation point with earlier waves of data collection for both age cohorts of children. This approach can increase statistical precision relative to annual impact analyses by taking advantage of the “time within student” portion of the data structure. Addition of a new time point, fully two years beyond any previously available (third grade rather than first grade) will greatly extend what the analysis can say about the individual and average growth trajectories of children as they move through elementary school and how access to Head Start at a younger age may have affected those trajectories

Repeated-measures longitudinal analysis will be conducted using hierarchical linear modeling (HLM) (Raudenbush and Bryk, 2002) that models individual children’s growth curves across time points nested within students, as well as students nested within the centers of initial random assignment. The application of HLM can be illustrated by considering a simple two-level model of time within students (for clarity of explanation a two-level model is used as an illustration). Suppose the outcome variable Y is observed for individual child “ i ” across multiple waves of data collection indexed by “ t ,” giving Y_{it} . This outcome is influenced by fixed background characteristics, W_i (e.g., child gender) measured at baseline, and another time-invariant factor, assignment to the Head Start treatment group, Z_i . Outcomes are also related to a “time varying” covariate, the precise time of data collection for wave t , T_{it} .

In this formulation, each child has his or her own growth “trajectory” that can be expressed in terms of a hierarchical model (Jo and Muthen, 2003) consisting of two levels:

- **Level 1: Time Within Students**, describes the different time points of data collection for a given child, starting with $t = 0$ for baseline measures collected in fall 2002):

$$Y_{it} = \pi_{0i} + \pi_{1i} T_{it} + e_{it}, \text{ where}$$

Y_{it} = outcome for individual i when measured in wave t ,

T_{it} = time of data collection for individual i in wave t ,

π_{0i} = growth curve intercept for individual i ,

π_{1i} = growth curve slope parameter indicating how outcome changes with time for individual i , and

e_{it} = random error term attributable to sampling variation.

- **Level 2: Between Students**, describes the way individual child characteristics relate to the level and growth rate of their outcomes:

$$\pi_{0i} = \beta_{00} + \beta_{01} W_i + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11} W_i + \beta_{12} Z_i + r_{1i}, \text{ where}$$

π_{0i}, π_{1i} , are random intercepts and slopes for growth curves, varying by individual,
 W_i = time-invariant covariate for child i (e.g. gender),
 Z_i = a binary treatment variable with $Z_i=0$ if student i is assigned to the control,
group and $Z_i = 1$ if student i is assigned to the treatment group, and
 r_{0i}, r_{1i} are random error terms attributable to sampling variation.

Note that if the formulas for the π factors from these equations are plugged into Equation (1), the treatment dummy variable Z_i and the time invariant covariates W_i once again influence the outcome Y_{it} as in a single-equation regression model proposed for the annual cross-sectional analysis. But here the individual children's growth trajectories are expressed in terms of their own personal intercepts (π_{0i}) and slopes (π_{1i}) and variation in these terms is related to the time-invariant background characteristics of those children, W_i , at Level 2 of the model. Assignment to the treatment or control group, Z_i , does not affect the intercept, π_{0i} , because initially the two samples are statistically indistinguishable; however, it does affect the growth rate of the outcome, π_{1i} .

A key parameter in this model is β_{12} , which indicates the impact of assignment to treatment (i.e., access to Head Start) on π_{1i} , the growth in Y over time. Though the dependent variable in the equation containing β_{12} is not observable, β_{12} itself (and all the other right-hand-side coefficients) can be estimated as facets of the latent growth trajectory of an individual child. The impact of treatment on Y at any point in time, which grows linearly with time T , can be calculated for child i by plugging the formula for π_{1i} into Equation (1) to get $\beta_{12}T_{it}Z_i$. This means that the outcome at time T_{it} is higher for child i when assigned to the treatment group ($Z_i=1$) than when assigned to the control group ($Z_i=0$), by an amount equal to $\beta_{12}T_{it}$. Just as in a single-level cross-sectional regression, this impact estimate is controlled for the value of the baseline covariate, W_i (which also carries through into Equation (1) when the formulas for the π coefficients are substituted), thereby increasing the statistical precision of the impact estimate. In addition, any time-varying covariates now included in the analysis, X_{it} , also help to explain variation in outcomes across children and over time, further increasing statistical precision.

Longitudinal analyses can also be used to address a number of research questions about the potential long-term outcomes associated with attending Head Start as well as questions related to how children's outcomes may be moderated by various experiences. Individual growth modeling provides one strategy for conducting these longitudinal analyses.

Variation In Impacts

In addition to estimating overall national impacts for all Head Start children, we will also assess the extent to which impacts at the end of third grade vary across different groups of children, parents, and families. This requirement responds to the second broad research question recommended by the National Advisory Panel: “*Under what circumstances does Head Start achieve the greatest impact? What works for which children? What Head Start services are most related to impact?*” These analyses will be conducted for the subgroups identified by OPRE as important within the context of Head Start participation and during the first two years of elementary school and used in the Final Report from the original evaluation:

Child:

- Age (3- vs. 4-year old cohorts);
- Presence of special needs;
- For cognitive and social/emotional outcomes only, whether child exhibits “low” academic ability at baseline;
- Language in which child was assessed for the baseline assessment in fall 2002.

Parent:

- Parent/caregiver reported depression;
- Biological mother/caregiver race/ethnicity;
- Parental “risk” index, consisting of the number of the following characteristics that reportedly exist for the family:
 - Receipt of TANF or Food Stamps
 - Not having a high school diploma or a GED (neither parent in household has high school diploma or a GED))
 - Not employed nor in school (neither parent in household is employed or in school)
 - Single parent (biological mother/primary caregiver)
- Teen mother, i.e., biological mother was age 19 or younger when study child was born

Community:

- Urban location
- Community socio-economic status

Additional subgroups may be added if new policy focuses emerge or child developmental issues at older ages (i.e., through age 9) so dictate; we would anticipate working with OPRE to identify any such appropriate additions. As in the original study, to determine impact, it is essential that all subgroups be defined by characteristics that could not have been affected by assignment to the treatment or control groups and hence define subsets of the evaluation sample that are fully comparable in the two groups.

The estimation of subgroup impacts is easily accomplished by adding an interaction term to either the OLS/logistic/categorical regressions used for the annual cross-sectional analysis, or to the student-level model for those analyses done within the repeated measures framework. This term interacts a dummy variable for inclusion in the subgroup of interest with the random assignment treatment/control group indicator, Z . The parameter coefficient associated with the added interaction term represents the extent to which impacts are higher or lower for children in the particular subgroup flagged by the dummy variable.

Linkages Between Experiences and Outcomes – Mediating/Moderating Analysis

Additional analyses will be undertaken to identify the role that schools may play in mediating and moderating the impact of Head Start.” These are extremely important analyses and one of the key reasons for conducting a Third Grade Follow-up of the HSIS children.

The fundamental reason for conducting an experiment—over other types of evaluation designs—is that random assignment creates two or more groups of individuals who do not differ systematically in any way except for the experimental treatment (or in some studies different distinct treatments). As a consequence, any subsequently observed overall differences between the experimentally created groups that exceed the bounds of sampling error can reliably be attributed to the effect of the treatment condition, i.e., this difference is what we call the estimated “causal impact” of the treatment or program.

This is precisely what the impact analyses described up to this point seek to accomplish—directly establish the causal effect of access to Head Start on child and family outcomes. But, the new type of question stated above (i.e., role that schools play in mediating and moderating the impact of Head Start) presents new challenges to the analysis. These challenges arise because the school experiences that could mediate the impact of Head Start *may themselves be affected by assignment to treatment*. This means that students with a particular type of school experience cannot simply be plucked out of the treatment and control groups and treated as equivalent, as was possible for characteristics defined when children were initially randomly assigned (e.g., language of initial cognitive assessment) used to define subgroups in the earlier discussion of variations in impact.

The seminal work of Harvard statistician Donald Rubin has brought great clarity to the problem of causal inference when the groups of interest are themselves defined in part by the results of randomization, through his development of the idea of “potential outcomes” (Ruben, 2004). The underlying concept is that an individual study child “i” could have been “potentially” assigned to either the treatment group (where $Z_i=1$) or to the control group (where $Z_i=0$). For each study child, we collected data on an outcome Y_i (e.g., the Woodcock Johnson III Letter Word Identification test) at a specific time after assignment to treatment status, and therefore $Y_i(Z)$ is the value of the outcome Y for child “i” if he/she is assigned to condition Z where $Z = 1$ or 0 . According to Rubin, the causal effect of assignment to treatment on outcome Y is the difference in potential outcomes for each individual child:

$$Y_i(Z_i=1) - Y_i(Z_i=0),$$

Further, the estimate of the average Intent To Treat (ITT) impact is:

$$E(Y_i(Z_i=1) - Y_i(Z_i=0)).$$

However, we cannot observe potential outcomes for all individual children because each child can only exist in a single state, i.e., we can only observe them in the state they experience—the group or treatment status to which they were randomly assigned. That is, for those children who we assign to the experimental treatment group (T) we have their outcomes under the condition that they were given access to Head Start (but not their outcomes if they had been denied access to Head Start). Similarly for those children who were randomly assigned to the experimental control group (C) we have their outcomes under the condition that they were NOT given access to HS (but not the outcomes if they had been). In our typical analyses of program impact, we take the difference in the mean **observed**

outcomes for the **entire** two experimentally created groups. (Rubin, however, argues that in fact every individual has a “true” potential outcome in both states, whether we observe it or not.)

The intuition here is that if the treatment and control group are alike in all characteristics that relate to the outcome **before** treatment, any subsequent difference in treatment and control group means is an unbiased estimate of the average difference in potential outcomes. Another way of saying this is that the potential outcomes of the individual children are not statistically related to treatment group assignment (Raudenbush, 2003).

A key aspect of this formulation is that the comparison that generates causal effects must be on individuals from a single population. That is, a comparison of one set of study subjects with another when the two groups are not comparable would not generate causal effects. Take, for example, our proposed analysis of subgroup impacts described above. When children were randomly assigned to either the treatment or the control group, the entire groups were, on average, equivalent. And, there was also equivalence for any subgroups of children defined on characteristics that could not have been affected by the process of random assignment such as the child’s gender, race/ethnicity, age, etc. Moreover, we can explicitly identify these subgroups in both arms of the experiment—that is, we can identify all of the boys and girls in the treatment group and their matched children in the control group.

This ability to compare outcomes for subgroups defined at the time of random assignment (so called “exogenously” or externally created groups) is not, however, possible for groups of individuals that are defined on the basis of events or actions that occur **after** random assignment, referred to as “endogenously determined” subgroups (i.e., created within the experiment itself). This is because the events or actions that are observed after random assignment may have been affected by the experimental treatment to which an individual was assigned, thereby creating subgroups that are not comparable. In some cases, this problem is further exacerbated by the fact that the characteristic that defines a particular subgroup is applicable to (or observed in) only one of the randomly assigned groups. In this case, it is impossible to identify the corresponding subgroup in the other experimental group. One example of this type of problem is a common policy question about the impact of exposure (or dosage) where individuals in the treatment group may opt to experience different amounts of the treatment. In this example, one cannot know which individuals in the control group correspond to the observed “dosage” subgroups in the experimental treatment group because they did not have the opportunity to exhibit the behavior of interest.

Three options for the analysis of linkages between children’s early school experiences and impacts on child and parent outcomes are described below.

An Interpretive Approach

The “simplest” option for finding linkages between children’s early school experiences and impacts on child and parent outcomes can be used as either a stand alone method or to supplement formal structural analysis of this connection (see below). This interpretive approach simply looks at the pattern of separately estimated and reported impacts on child and parent outcomes, and compares these results to the separately estimated impacts on children’s experiences. The goal would be to see if the pattern of impacts on child/parent outcomes is consistent with the pattern of estimated impacts on experiences/services. For example, if impacts are found on children’s language and literacy outcomes, we would look to see if there are associated impacts on the nature and type of reading instruction that children receive in school. Such consistent patterns are not, of course, evidence of causal linkages because there are many potential paths by which a particular school readiness outcome can be impacted. Hence, the presence of a significant effect on the child experience indicator being examined does not rule out the possibility that Head Start changed other types of experiences and it was through those channels rather than the one under scrutiny that child outcomes were enhanced. Conversely, the absence of a statistically significant impact on the experience measure does not prove *lack of influence*, only a lack of the ability to detect whatever impact in that experience area may have occurred. So, while we cannot prove with certainty any particular causal linkage with this interpretive approach, it may at be least suggestive of the channels by which effects on children’s school experiences may be influencing the impact of access to Head Start on various developmental outcomes.

Leveraging the Experimental Design

This option builds on the strength of the experimental design by using internally valid estimates of program impact within study sites, where sites are defined as the centers (or center groups) where random assignment was conducted. The methodology was first developed as part of the Head Start Impact Study Research Design and Analysis Plan (2003) and later independently suggested by Bloom (2005) using multilevel modeling to capture both individual- and site-level factors that are related to the outcomes of interest. This method would extend either the annual cross-sectional analysis, or the repeated measures analysis described above, to include a new covariate which is defined as the **difference** in a particular experience measure (e.g., access to a “high quality” third grade teacher) between the treatment

and control group children who were randomly assigned at a particular center. The number of added variables would be equal to one less than the number of centers (or rather center groups as used in the sample selection process).

Underlying this model is the important assumption that **systematic treatment-control group differences in experiences** are what give rise to observed variation in site-level impacts on child and parent outcomes. In effect, this method uses the child-level data to estimate a “difference in experiences” statistic, $(\tau\hat{S}_{ps} - c\hat{S}_{ps})$, for each selected experience indicator, by individual study centers. For example, a model such as the one shown above can include treatment-control differences related to the extent to which randomly assigned children differ on a particular measure of their early school experiences including third grade experiences. The key requirement is that the particular experience measure must be definable for each individual child, something that should not be a problem once all children in the sample are attending school.

While this analysis gives strong answers to the questions about “what works” it is subject to the criticism of mis-specification bias because not all the factors that influence site-level impacts are likely to be included in the model, nor are those that are included necessarily well measured. Where these conditions are not met, the estimated influence of school experiences on child impacts may be biased; the occurrence and extent of this bias depends on whether the omitted or incompletely measured site-level factors correlate with the featured variable in the analysis, Head Start’s impact on child experiences. In other words, it may be that there is no systematic relationship between the centers where Head Start has a comparatively large impact on children’s experiences and centers with especially dynamic directors whose energy infuses her staff and increases school readiness impacts at that center in and of itself; if there is not, the model will have left out an important determinant of center level impacts but not one whose influence gets falsely attributed to differences in Head Start’s impact on child experiences across centers.

Growth Modeling Framework

By the time children reach third grade there are many community, family, home, and school experiences that may affect their cognitive, social-emotional, and health outcomes. Understanding the effects these experiences may have on outcomes and whether some of the outcomes are conditioned by varying experiences are important to address. These analyses move further away from the experimental design, but they provide the opportunity to assess whether other influences are at work behind or in place

of measured Head Start impacts. They are used to answer such questions as, “To what extent do the impacts of Head Start depend upon children’s early experiences and experiences in third grade?”

We provided a description of Hierarchical Linear Modeling (HLM) to provide answers to longitudinal growth questions about impacts. HLM can also be used to help answer questions that are non-experimental in nature. For example, research questions that examine the child and classroom factors that may moderate the effects of Head Start can be tested by extending these equations to include interaction terms between whether the child attended Head Start, child or classroom characteristics, and time (in the case where the analyses examine whether the effects of Head Start on children’s rates of development are moderated by child or classroom characteristics).

Latent growth analysis may also be applied to study the effects of attending Head Start when there are repeated assessments of children’s developmental outcomes, and AMOS, LISREL, M-plus or other structural equations modeling software may be used to conduct these analyses. In order to simplify the equations below, time and Head Start attendance are included as the only predictors of children’s developmental outcomes. Actual analyses of these data will also include various covariates related to child, family, and experiences within classrooms, including variables that change over time (e.g., family income, classroom experiences) and variables that are stable over time (e.g., gender, race).

Assuming a series of repeated measurements Y_{it} (minimum of three repeated measurements; i represents an individual, and t represents the order for measuring Y), the growth trajectory model for describing an individual’s growth as represented by this series of data takes the form (Level 1 or within-person model):

$$Y_{it} = \alpha_i + \beta_i \lambda_t + \varepsilon_i$$

where α_i represents the intercept of an individual’s growth trajectory (i.e., the initial status at Time 1), β_i represents the slope of the individual’s growth trajectory (i.e., the unit change in Y_i between two consecutive measurements), λ_t represents ordered time points at which the measurement is taken, and finally, ε_i represents the modeling residual for an individual. Because the intercept (α_i) and the slope (β_i) are random variables that vary across children, these individual model parameters can be represented by the group mean intercept (μ_α) which applies to all children and a group mean slope (μ_β), plus individual variation ($\zeta_{\alpha i}, \zeta_{\beta i}$) as follows (Level 2 or between-person model):

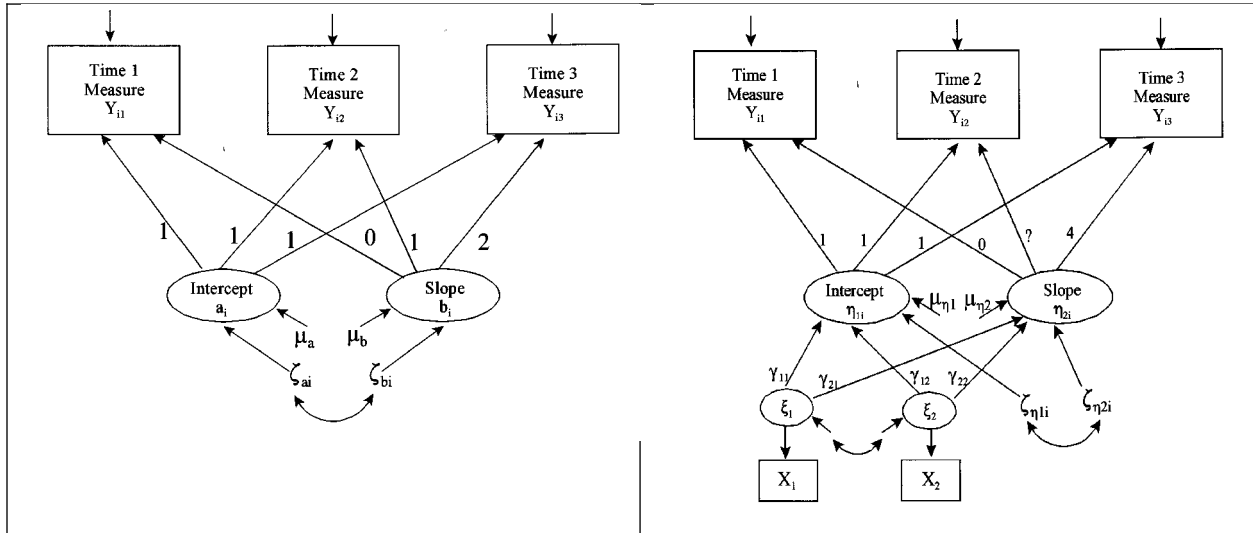
$$\begin{aligned}\alpha_i &= \mu_\alpha + \zeta_{\alpha i} \\ \beta_i &= \mu_\beta + \zeta_{\beta i}\end{aligned}$$

Level 2 assumes that individual trajectory parameters (α_i , β_i) are only a function of the group mean trajectory parameters plus individual variation, and no other predictors are involved in the model to account for the across-child variation of the trajectory parameters. As a result, the Level 2 model presented above is often called an unconditional model (e.g., Curran, 2000). In situations where research calls for testing for predictors (X_i) that may account for potential systematic variation of the individual trajectory parameters, such as in the proposed analysis of the effects of Head Start, a conditional Level 2 (between-person) model can be constructed as follows:

$$\begin{aligned}\alpha_i &= \mu_\alpha + \gamma_1 X_{1i} + \zeta_{\alpha i} \\ \beta_i &= \mu_\beta + \gamma_2 X_{1i} + \zeta_{\beta i}\end{aligned}$$

where, X_1 is Head Start attendance that may account for potential systematic variation in the trajectory parameters (intercept α and slope β), and γ_1 and γ_2 represent the predictive effects of attending Head Start on the trajectory intercept (α) and slope (β). Within the framework of structural equation modeling, the unconditional latent growth model for both within-person and between-person components can be represented in Figure 1 (for a situation of four repeated measures). In this model, the coefficients from the intercept to the repeated measures is constant, while λ_t (ordered time at which one of the repeated measurements is taken) takes the values of 0 (initial measurement), 1 (second measurement), and 2 (third measurement). The values of λ_t , however, can be specified to reflect the time actually passed between each consecutive measurement. Further, hypotheses of curvilinear or unequal growth pattern can be empirically evaluated in the LGA model as generally represented in Figure 1.

Figure 1. Unconditional and Conditional Latent Growth Models



For the conditional between-person LGA model for testing predictors that may account for potential systematic variation in the growth trajectory parameters (intercept and slope), the model can be represented by the structural equation model on the right in Figure 1. In this model, X_1 is whether the child attended Head Start and X_2 represents a characteristic of the child or family (more can be incorporated into the model as called for in a specific research situation), which are included to test if they explain the variation of the individual growth trajectory parameters (i.e., intercept and slope). The path coefficients (γ_1 to γ_2) from these explanatory variables to the intercept and slope represent the effects of these variables on the growth trajectories.

For example, we may ask if Head Start attendance (X_1) contributes to the difference in the initial status (i.e., the intercept) of the growth trajectory, or if it accelerates the growth (i.e., steeper slope for the trajectory). The model described in Figure 1 can easily be expanded to accommodate more explanatory variables similar to X_1 and X_2 in the model. In addition, research questions that examine the child and classroom factors that may moderate the effects of Head Start can be tested in this framework, by extending Level 2 equations above to include interaction terms between whether the child attended Head Start, child or classroom characteristics, and time (in the case where the analyses examine whether the effects of Head Start on children’s rates of development are moderated by child or classroom characteristics).

It should be noted that the growth trajectory over a time span does not have to be linear. If there are enough time points in the data collection process, different curve structure of the growth trajectory can be tested, such as linear, quadratic, or some other shape of trajectory. For analyses in which there are more than three observations of children’s academic competencies, different growth curve structures will be examined to identify the growth function that best fits the patterns in which children develop particular competencies over time.

Tabulations

The most detailed information on estimated program impacts (both intent to treat (ITT) and impacts on the treated (IOT)) will be provided in a series of tables that will include the following information:

- **Third grade outcomes by domain and age cohort (3- and 4-year-olds) including:**
 - Treatment and control group means (with standard deviations),
 - Difference in means (with standard deviation),
 - 95-percent confidence intervals around the estimated mean difference, and
 - Associated p-values (i.e., indicating statistical significance).
- **Estimated impacts for the third grade follow-up**, also by cohort and organized by domain to coincide with the chapters in the main report. Tables will include the following:
 - **Mean differences:** treatment and control group means (with standard deviations), difference in means (with standard deviation), 95-percent confidence intervals around the estimated mean difference, and associated p-values (i.e., indicating statistical significance).
 - **Regression-based impact estimates:** estimated impact (with standard deviation), 95-percent confidence interval, p-value, estimated effect size (with standard deviation), and 95-percent confidence interval for the estimated effect size.

These tables will be very lengthy and will be filled with a great deal of data. Consequently, to make the “story” more understandable to the reader the main body of the final report will focus on the following:

- Graphical displays that present estimated impacts, expressed as “effect sizes,” at the end of the preschool period—representing the impact on school readiness—and at the end of the observed school period (i.e., end of third grade). The graphs will, as shown, be organized by domain (matching the report chapters) and will show both age cohorts together for each outcome measure. The 95-percent confidence intervals and statistical significance will also be provided.
- Where growth curve analysis is used within the cognitive domain, we will also provide growth trajectories illustrating the average development of the treatment and control group over time, for each of the three selected cognitive outcomes.

Publications

In addition to the Final Study Report, we will produce at least one article for a peer-reviewed journal, at least one Issue Brief, and three conference presentations including a presentation at the Head Start Research Conference. Table 7 provides the timelines for these publications and presentations.

Time Schedule.

Table 7 summarizes the deliverables for the Third Grade Follow-Up Study.

A.17 Approval to Not Display the OMB Expiration Date

We do not plan to ask for an exception to the OMB rule that the expiration date be printed on all survey documents.

A.18 Exception to the Certification Statement

There are no requested exceptions to the certification statement in Item 19.

Table 7. Project Deliverables and Delivery Dates

Item	Description	Task	Due Date
01	List of Proposed Consultants	Task 2	10 Days
02	Schedule of Consultant Meeting	Task 2	6 Weeks
03	Topics for Agenda (Consultant Meeting)	Task 2	6 Weeks Prior to Consultant Meeting
04	Minutes of Consultant Meeting	Task 2	2 Weeks after Consultant Meeting
05	Draft Tracking & Attrition Plan	Task 4	6 Weeks
06	Final Tracking & Attrition Plan	Task 4	10 Weeks
07	Draft of Detailed Data Collection Plan	Task 3	6 Weeks
08	Final Detailed Data Collection Plan	Task 3	10 Weeks
09	Reports of Progress Recruiting Families	Task 4	Bi-weekly Throughout Task 4 & Task 5
10	Outline for Training Program and Plan for Monitoring Quality of Data Collection	Task 5	8 Weeks Prior to 1 st Training Session
11	Draft Training Materials and Plan for Monitoring Quality of Data Collection	Task 5	4 Weeks Prior to 1 st Training Session
12	Final Training Materials	Task 5	5 Working Days Prior to being submitted to training participants
13	Summary of Quality Control Site Visits	Task 5	Within 2 Weeks of the conclusion of each visit
14	Draft List of Technical Notes to be Written	Task 6	28 Weeks
15	Final List of Technical Notes to be Written	Task 6	32 Weeks
16	Data Deliverables to FPO	Task 6	4 Months Prior to the Contract End Date
17	Final Data Sets on Head Start Archive	Task 6	1 Month Prior to the Contract End Date
18	Progress Reports	Task 7	Monthly
19	Draft Presentations for FPO Review	Task 7b	3 Weeks Prior to Each Presentation
20	Draft Outline of Study Report	Task 7b	40 Weeks Prior to Contract End Date
21	Draft Study Report	Task 7b	28 Weeks Prior to Contract End Date
22	Final Camera-Ready Copy, and Electronic Copy of the Study Report	Task 7b	2 Weeks Prior to Contract End Date
23	Draft Outline for Journal Article	Task 7b	40 Weeks Prior to Contract End Date
24	Draft Journal Article	Task 7b	28 Weeks Prior to Project End Date
25	Final Journal Article (Ready for Submission)	Task 7b	2 Weeks Prior to Contract End Date
26	Draft Outline for Issue Brief(s)	Task 7b	40 Weeks Prior to Contract End Date

Table 7. Project Deliverables and Delivery Dates (continued)

Item	Description	Task	Due Date
27	Draft Issue Brief(s)	Task 7b	28 Weeks Prior to Project End Date
28	Final Camera-Ready Copy, and Electronic Copy of the Issue Brief(s)	Task 7b	2 Weeks Prior to Project End Date
29	Project Work Plan	Task 7b	4 Weeks

B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS

B.1 Potential Respondent Universe

The potential respondent universe for the Third Grade Follow-Up Study is the same as reported for the Head Start Impact Study. The sampling universe for the Head Start Impact Study is all Head Start grantees/delegate agencies operating in all 50 states, the Commonwealth of Puerto Rico, or the district of Columbia that do not specifically target the following special populations: grantees/delegate agencies serving migrant children; Head Start programs operated by Tribal organizations; children enrolled in Early Head Start; new grantees/delegate agencies (i.e., grantees/delegate agencies that are in operation for less than two years); and programs involved in FACES 2000.

The starting point for creating this initial population of Head Start grantees/delegate agencies was the 1999-2000 Program Information Report (PIR) database maintained by ACYF. Migrant and Tribal Organization grantees/delegate agencies can be readily identified from this database, and “new” programs were identified as those grantees/delegate agencies that were listed in the 1999-2000 PIR but which were **not** listed in the 1998-1999 PIR (i.e., eliminating grantees/delegate agencies that were in operation for approximately less than two years). Early Head Start children were identified and excluded once the sample of grantees/delegate agencies has been selected.

B.2 Description of Sampling and Information Procedures

Sampling Procedures

In fall 2002, a sample of 4,667 children was randomly selected off the applicant lists to a sample of Head Start centers. This sample was randomly assigned within centers (in a few cases, within programs) to either a treatment or control group in a sample size ratio of 1.5. The fall 2002 sample sizes are given in Table 8 below, with the anticipated grade 3 sample sizes based on the response rates obtained at the most recent spring data collection to date, which was in 2005 for the age 4 cohort and in spring 2006 for the age 3 cohort. An attempt will be made to assess all children sampled in fall 2002 who reach third grade in either spring 2007 or 2008, and to interview their parent(s).

Table 8. Expected Sample Sizes in the Third Grade Follow-Up Study

	Fall 2002 sample size	Spring 2005/2006 response rates	Expected sample size grade 3
Parent Interview			
Head Start Treatment Group	2,783		2,227
age 3	1,530	83%	1,270
age 4	1,253	82%	1,027
Control Group	1,884		1,413
age 3	1,029	75%	772
age 4	855	75%	641
Combined Treatment & Control	4,667		3,710
Child Assessment			
Head Start Treatment Group	2,783		2,226
age 3	1,530	80%	1,224
age 4	1,253	80%	1,002
Control Group	1,884		1,357
age 3	1,029	72%	741
age 4	855	74%	616
Combined Treatment & Control	4,667		3,583

The Head Start Impact Study Sampling Plan is included in Appendix E.

Weights, Nonresponse Adjustment, and Variance Estimation

Sampling weights will be calculated for each child to allow estimates based on the sample to represent the population of newly entering Head Start applicants in the third grade. Each child will be assigned a base weight that reflects their overall probability of selection in the HSIS, including the sampling of geographic areas used as primary sampling units (PSUs), Head Start grantees/delegate agencies, centers, and the randomization of children to treatment and controls off each center’s applicant list. These will be the same base weights that were developed for the HSIS cohort sampled in 2002 because the third grade cohort will be a followup of the same children. The base weights will be adjusted for nonresponse to the child assessments, the parent interview and other instruments (e.g., teacher interview, classroom observations) for the data collection in spring 2007 and 2008, when the original cohort age 3- and 4-year-olds has reached the third grade.

The nonresponse adjustment helps control for nonresponse bias by compensating for different data collection response rates across various demographic and geographic groups of children. This is due to the fact that the nonresponse adjustment factor is calculated within nonresponse adjustment

cells formed by the demographic and geographic variables that are most correlated with response rates, as determined by a separate response propensity analysis. The nonresponse adjustment factor spreads the weight of the nonresponding children over the responding children within each cell, so that they represent not only the children who were not sampled, but also the nonresponding sampled children. This maintains the same mix of the sample across cells as would have been present had there been no nonresponse.

Estimates obtained from the Third Grade Follow-Up study will differ from the true population parameters because they are based on a randomly chosen subset of the population, rather than on a complete census of all newly entering Head Start applicants. This type of error is known as sampling error or variance. The differences between the estimates and the true population values can also be caused by nonsampling error. Nonsampling errors can result from many causes, such as measurement error, nonresponse, sampling frame errors, respondent error, and differences among interviewers. In general, the magnitude of nonsampling error is difficult to assess from the sample. The precision of an estimate is measured by the standard error (defined as the square root of the variance). The calculation of the standard error must reflect not only the sample size on which the estimate is based, but the manner in which the sample was drawn. Otherwise, the standard errors can be misleading and result in incorrect confidence intervals and p-values in hypothesis testing. The HSIS sample involved stratification, clustering, and unequal probabilities of selection, all of which must be reflected in the standard error estimate.

Two commonly used variance estimation methods for complex surveys involving multi-stage sampling are replication and linearization (Wolter, 1985). Replication methods work by dividing the sample into subsample replicates that mirror the design of the sample. A weight is calculated for each replicate using the same procedures as for the full-sample weight. This produces a set of replicate weights for each sampled child. To calculate the standard error of a survey estimate, the estimate is first calculated for each replicate using the replicate weight and the same form of estimator as for the full sample. The variation among the replicates is then used to estimate the variance for the full sample estimate. In the linearization approach, a nonlinear estimator is approximated by a linear function and a formula derived for the variance of the linear approximation. Replication has the advantage that it can reflect the different features of the weighting and estimation by simply repeating all steps separately for each replicate. For linearization, a specific formula is needed for each estimator, and the formula will differ depending on the type of estimator and sample design. On the other hand, finite population correction factors are often easier to account for using linearization estimators. However, for linear estimators, or nonlinear

estimators that are formed by combinations of linear functions, replication variance estimators are often little different numerically from linearization variance estimators.

To facilitate standard error estimation, we will provide a set of jackknife replicate weights and the sample design variables necessary to calculate standard errors via linearization, for use with such software packages as Wesvar, SUDAAN, Stata, and the new survey procedures (proc surveymeans, proc surveyfreq, proc surveyreg, proc surveylogistic) in SAS version 9. We will also provide the necessary weights to run HLM v6 for the estimation of multi-level models.

Power of Tests

Power measures the ability of a statistical test to detect specified differences between two or more groups. By ‘detect’ we mean show statistical significance. Power calculations are generally performed in the course of designing a study to determine the sample sizes required to detect interesting or “important” differences between groups. In this section, the results of a power analysis for the detection of a Head Start impact are given based on the expected sample sizes of grade 3 children in spring 2007 and 2008. The grade 3 children will be those followed from the 3-and 4- year-old cohorts sampled in 2002 for the HSIS. The minimum effect size detectable with a power of .80 by a t-test of no significant difference between treatment and control groups is given for the age 3 and 4 cohorts for a few important outcomes from the child assessments and the parent interview.

There are no universal standards for how much power is “good enough”. In some contexts a power as low as 50% is considered adequate; in others a power of 95% or higher may be required. In the absence of specific requirements, a power of 80% is generally considered good. The power of statistical tests for surveys depends on a number of factors, including:

- The treatment and control sample sizes.
- The precision of the estimates, including the effects of the survey design (e.g., stratification, clustering, and weighting) on the target variables, as measured by the design effect.
- The size of the difference to be detected, sometimes measured as the *effect size*.
- The functional form of the estimator, e.g. means, proportions, regression coefficients. For a proportion, the standard error is a function of the proportion. Proportions of 50% have the lowest precision relative to proportions close to 0 or 100% based on the same sample size.

- The distribution of the variables whose parameters are being estimated, e.g., normal, lognormal, binomial, etc., including the mean and population standard deviation of the distribution.

Power Analysis Results

The minimum effect size detectable at power=.8 for a two-tailed t-test of the hypothesis of no difference in the treatment and control group means for four continuous child assessment outcomes is given in Tables 2-9a and 2-9b below. The type I error of the t-test is .05. All minimum detectable effects are based on the effective sample sizes, to reflect the HSIS sample design. (The effective sample size is equal to the nominal sample size divided by the design effect for the estimate. The design effect is calculated as the variance under the sample design divided by the variance under simple random sampling.) The mean and population standard deviation of the distribution for each outcome variable and the design effects were estimated from spring 2003 NHIS data, based on children with completed spring 2003 assessments. The minimum detectable effect size is calculated as the difference in treatment and control group means divided by the population standard deviation. The table shows that for most continuous child assessment outcomes the MDE is about .20-.22

For regression-adjusted estimates of treatment vs. control group differences in mean outcomes, in which the child outcome is modeled as a function of school, classroom, teacher, child and parent characteristics and a treatment group indicator, the power to detect differences between groups is improved if the covariates in the model have good explanatory for the outcome. In fact the variance of the estimates is reduced by a factor of $1-R^2$ (Cochran, 1983, Eqn 6.9.5), where R^2 is the squared multiple correlation coefficient between the outcome and the covariates.

Minimum detectable differences for binary outcomes are given in Table 9a. The population proportions and design effects were estimated using spring 2003 NHIS data for two binary outcomes, based on children with a completed spring 2003 parent interview. The spanking outcome (whether the child was spanked in the last week) was chosen because its estimated proportion for the control group is about 50%, where precision is lowest for a proportion. By contrast the “number of times per week child read to” outcome is a categorical variable with only 6 to 7% of control parents reporting reading just once a week to their child, as opposed to more than once. This serves as an example of an estimated proportion whose precision is expected to be higher since it is closer to 0%, thus the minimum detectable difference for a power of .8 should be smaller than for the proportion of 50%. Table 9b shows that the minimum detectable difference for a proportion in the 40 to 50% range is 9 to 10 percentage points, taking the

design effect into account, while smaller differences of 5 to 7 percentage points would be detectable for proportions closer to 0 or 100%.

Table 9a. Minimum Detectable Effect Sizes with Power=.8 for the Third Grade Follow-Up Sample

Outcome	Control sample size in grade 3	Treatment sample size in grade 3	Deff for control mean	Deff for treatment mean	Effective sample size for control group	Effective sample size for treatment group	Pop. mean	Pop. std. dev.	Minimum detectable effect size for power=.8
PPVT									
Age 3	741	1,224	1.81	4.3	418	287	250.0	35	.2
Age 4	616	1,002	2.18	2.6	289	384	291.3	40	.2
Total Child Behavior Problems									
Age 3	741	1,224	2.3	2.5	323	488	6.3	3.6	0.2
Age 4	616	1,002	2.5	1.8	255	573	5.8	3.7	0.2
Woodcock Johnson Spelling									
Age 3	741	1,224	2.4	2.4	311	523	91.7	11.8	0.2
Age 4	616	1,002	2.2	4.01	293	253	89.5	13.7	0.2
Woodcock Johnson Word									
Age 3	741	1,224	1.9	3.5	402	358	89.0	19	0.20
Age 4	616	1,002	3.1	5.9	202	172	90.4	15	0.3

Table 9b. Minimum Detectable Differences in Proportions with Power=.8 for the Third Grade Follow-Up Sample, Based on Spring 2003 HSIS Data

Outcome	Control sample size in grade 3	Treatment sample size in grade 3	Deff for control proportion	Deff for treatment proportion	Effective sample size for control group	Effective sample size for treatment group	Pop. Proportion	Minimum detectable difference for power = .8
Spanked in Last Week?								
Age 3	772	1,270	1.9	2.7	406	470	50%	9%
Age 4	641	1,027	2.2	2.6	303	405	40%	11%
Read to Only Once in Last Week?								
Age 3	772	1,270	2.0	2.4	386	529	7%	6%
Age 4	641	1,027	2.5	2.9	267	363	10%	8%

Data Collection Procedures

Westat will continue to use the managerial and field staff employed on the HSIS to implement the Third Grade Follow-Up Study. This arrangement capitalizes on the extensive experience, trust, and rapport achieved by our staff to form positive long-term relationships with the families and children selected for the study over the years to obtain cooperation and maintain participation. Many study staff have worked with the same families since the beginning of the study, and have built trusted and comfortable relationships with them.

Retention rates for supervisors are very high and generally are in the 90 to 100 percent range from one spring to another with the retention rates for interviewer/assessors running about 70 percent. Westat will train and use highly experienced, skilled field teams consisting of site coordinators and interviewer/assessors under the supervision of the HSIS's operations director. The operations director will implement the data collection plan and ensure monitoring and quality control tasks are conducted in an efficient, organized, and timely manner to continue the collection of high quality data and achieve the high response rates consistently realized on the HSIS.

Our site coordinators will continue to serve in their pivotal role as the primary local contact; assisting with securing participation of schools; enlisting cooperation and maintaining participation of respondents including parents, children, teachers, and principals and other administrators; coordinating all data collection activities in the site; tracking study participants; managing field staff; and ensuring quality control. They will report to the central office operations director and her staff and will manage a measurement team consisting of Westat field interviewers/assessors. The size of the team and the time allocated will be adjusted according to the number of cases and their location within the site. These teams will be responsible for scheduled data collection activities for each wave, including conducting in-person parent interviews and telephone or in-person parent updates, administering one-on-one in-person child assessments, and distributing and collecting teacher instruments (teacher surveys and teacher's child report forms [TCRs]) and principal surveys that contribute to assessments of individual children and program quality. The following briefly describes each of these data collection activities.

Interviews with Parents/Primary Caregivers. The main parent interview, no more than 1 hour in length, will be conducted in spring 2007 for the 4-year-old cohort and in

spring 2008 for the 3-year-old cohort. These interviews will be designed to gather information on key constructs such as parenting styles and practices, engagement, advocacy, well-being, education, employment, monitoring and supervision, etc. In addition, tracking updates will be conducted each fall and then for the 3-year-old cohort in spring 2007 and the 4-year old cohort in spring 2008 to keep in touch with families and verify critical tracking and setting information. In these brief, approximately 10 minute telephone interviews, field staff gather critical contact information as well as the child's current school classroom setting information to inform where data will need to be collected in future rounds. The combination of these two types of interviews separated by short intervals keeps parents actively involved in the study and enables tracing to occur in a timely fashion to maximize the likelihood of locating the family.

Advance letters telling the parent/primary caregivers that the interviewer/assessors will be calling to schedule an appointment for the interview will be mailed out in early March. The site coordinator and the interviewer/assessors will make arrangements for the parent interviews. Parent interviews will be conducted in the language of the respondents with English and Spanish versions available. If we are unable to hire a field interviewer to cover all language possibilities, we will enlist the aid of an interpreter, but in all cases, that person would need to sign the Westat Data Collector Code of Conduct and Assurance of **Privacy**. Once again, we will interview one primary caregiver of the child—preferably and in most cases—the same primary caregiver interviewed throughout the course of the HSIS. Most often this is the mother, if she is present in the home, or the biological father, or, failing that, a knowledgeable substitute

We recommend that the spring parent interviews primarily be conducted in the respondents' homes. This is the most convenient place for the majority of respondents. However, alternative locations such as public libraries may be used for respondents who do not want the interview conducted in their homes. The interviewer will use a laptop computer to conduct the interview.

Child Assessment. Children in the 4-year-old cohort will be assessed for the Third Grade Follow-Up Study individually in spring 2007 while those in the 3-year-old cohort will be assessed in spring 2008 when the vast majority will be in third grade. We plan to assess all children with their cohort group including those who have not advanced to third grade with their peers.

The child assessment will focus on the areas of reading (language and literacy), mathematics, and executive functioning. In addition, children will be asked to respond to a self-

report instrument that includes items about school, their attitudes, motivation, relationships and behavior. The total battery will take about 1 hour to administer. We will assess each child using the following instruments: Early Childhood Longitudinal Study-Kindergarten (ECLS-K) reading assessment, the Letter-Word Identification, Applied Problems, and Calculation tests from the Woodcock-Johnson III Tests of Achievement, the Numbers Reversed and Auditory Working Memory tests from the Woodcock-Johnson III Tests of Cognitive Abilities, and the ECLS-K Self-Description Questionnaire. In order to continue to measure Spanish language skills, children who were classified as bilingual in the HSIS will also be administered the Bateria Woodcock-Muñoz Identificación de letras y palabras. Spanish speaking children (Puerto Rico) will be administered the Identificación de letras y palabras, Comprensión de textos, Problemas aplicados, and Cálculo tests from the Bateria Woodcock-Munoz Tests of Achievement, the Inversión de números test from the Bateria Woodcock-Muñoz Tests of Cognitive Abilities, and translated versions of the Auditory Working Memory test from the Woodcock-Johnson III Tests of Cognitive Abilities and ECLS-K Self-Description Questionnaire. Following the assessment, the assessors will continue to complete a set of ratings that describes the child's behavior and attitude during the assessment and to record any special circumstances or concerns that may be important in interpreting the assessment results.

The child assessment will be administered in the child's home by a Westat trained field assessor. The assessor will enter the child's responses into a laptop computer using the Computer-Assisted Personal Interview (CAPI) technology. The SDQ will be designed as an Audio Computer-Assisted Self Interview (ACASI). These technologies were discussed in Section A.3.

Teachers. For the Third Grade Follow-Up Study, we will continue to have the study children's teacher complete the paper and pencil self-administered instrument we fielded for HSIS which includes the general teacher survey items and the child specific teacher/child report form items. Teachers of the 4-year-old cohort will complete the instrument in spring 2007, while teachers of the 3-year-old cohort will complete them in spring 2008.

The teacher survey once again will be limited to no more than 30 minutes to complete. The survey will include demographic and biographical information, such as education and years of experience, characteristics of the school and classroom, classroom management, and belief scales to assess teacher attitudes about working with and teaching children. The TCR, an individual child rating form, is included in the teacher survey. The individual child questions will request information on the child's school background, academic skills, school

accomplishments, health and developmental conditions, teacher child relationship, and child behavior. Once again, we will translate these instruments with the exception of standardized scales into Spanish for Spanish-speaking teachers.

Interviewer/assessors and site coordinators will be responsible for distributing and collecting the teacher surveys and TCRs during the first month of data collection. Our experience has demonstrated that in-person follow-up is most effective in getting high compliance from schools and teachers.

Principals. The school principal is another source of data for school demographic characteristics and quality indicators for the school, teachers, and classrooms. The principal survey will be distributed and collected by the interviewers/assessors and site coordinators along with the teacher surveys and TCRs. The survey is self-administered and will take about 20 minutes to complete. Topics to be covered in the survey include: school characteristics, school facilities, leadership and governance, school policies and practices, educational goals, professional development opportunities, and demographic information on the principal. If necessary, the principal will be translated into Spanish.

Secondary Data. We will use information from secondary sources such as the Common Core of Data (CCD) or the school or school district website, to track a school's record with respect to such issues as attendance, disciplinary issues, immunizations of children, average test scores, number of children receiving free or reduced price lunch, school improvement status, and teacher/student ratio. This information will be collected by Westat staff and will provide measures of school quality.

B.3 Methods to Maximize Response

Locating families for followup interviews and assessments is a necessary but not sufficient condition to minimize sample attrition. Once families are located, they need to agree to participate in the followup data collection. We will employ several strategies to minimize sample attrition and obtain a high response rate.

Minimizing attrition requires continuing to engage families who have participated in the past, convincing families who have not participated in recent rounds to rejoin the study, and

finding a strategy to assess children in cases where the family has moved more than 70 miles away from the study area. Each of these three components required to minimize attrition are discussed below.

Engaging Current Participants. To keep respondents interested in continuing their participation, we will emphasize that the Third Grade Follow-Up Study is new and important. We will develop a new brochure that explains the study to the parents/primary caregivers and details the importance of their and their child’s participation. Advance letters also will be mailed out with the brochure that explains the study’s significance.

Incentives can motivate parents to participate by demonstrating the extent to which we value their time and insights. We will provide an incentive of \$30 for the Third Grade Follow-Up Study to acknowledge the value of their participation in the new study. Each child will receive a non-cash incentive not to exceed \$5.

Converting Families Who Have Refused in the Past. Westat has established procedures for converting refusals and working with difficult cases. A refusal conversion letter is used in an attempt to demonstrate to the families the benefit of the study. A “sorry I missed you” card is left at a house if no contact is made with the parent/primary caregiver. If a parent/primary caregiver participated in previous rounds of data collection, but is now suddenly reluctant, the interviewer/assessor is trained to work with the parent to determine the reason for the refusal. The interviewer/assessor will work with the parent/primary caregiver to ensure them that the interview will be conducted at their convenience and as efficiently as possible. For some reluctant participants, assigning the case to another interviewer/assessor may encourage the parent to participate. In this instance, the newly assigned interviewer/assessor is trained to acknowledge any earlier contact with the parent. For example, the conversation might begin with, “I know we have called you before. I just wanted you to know how important you are to our research.” Interviewers/assessors are trained to determine the reasons for non-participation such as fear of the government, loss of benefits, language problems, too many research or marketing requests at dinner time, etc. and have tools available to assist them in converting the respondents. Tools include waiting a week or so before contacting the parent/primary caregiver, calling at different times during the day, sending a refusal conversion letter, or visiting the home (always when there is no telephone). The best way to avoid problems is to thoroughly document cases and review each case carefully. Interviewers/assessors maintain a record of contacts for each case and thoroughly document the dates, times and outcomes of any contact with a

respondent or their household. Westat will implement these procedures for every case to insure the highest response rates possible for the study.

Through every data collection of the original HSIS, the interviewers tried to contact each family who was randomly assigned to the study. The only exceptions were cases that have been determined to be “hard refusals.” These were individuals who told us—in no uncertain terms—that they do not wish to participate in the study and did not want to hear from a representative of the HSIS again. While respondents can be encouraged and often persuaded to participate, participation *is* voluntary. It is important for the integrity of the study that respondents participate at will. Because the Third Grade Follow-Up Study is a new study, we will try to re-engage all of these families for the new effort.

Interviewing and Assessing Movers. We have a number of families who have moved more than 70 miles from the site from which they were randomly assigned and are more than 70 miles from any other site where we have interviewers working. We have continued to track these families and have completed the parent interview over the telephone. Unfortunately, the child assessment requires an in-person administration so it could not be completed in these situations. In spring 2006 we completed 69 3-year-old cohort parent interviews over the telephone where we could not assess the child because the family had moved. If we were able to complete these child assessments, we could increase our response rate on the child assessment by 2.7 percent.

We plan to use travelers to go to the areas where these families have moved to conduct the child assessment and parent interview in-person. We feel that it is one method that will be effective in minimizing attrition and keeping our response rates high. These families have indicated a willingness to participate in the study. With time, movers become more prevalent in the sample. Because the data from the child assessment is so valuable, interviewing just the parent is a less than optimal tactic for measuring the impact of Head Start on third grade students.

By using the strategies outlined above to reconnect with hard refusals, conduct child assessments with movers, and offer an increased incentive to parents and children, we believe that we will keep families from leaving the sample and will be able to re-engage families who have not participated in more recent rounds.

Elementary School Principals and Teachers. In advance of spring data collection, principals with study children enrolled will be mailed an introductory letter to describe spring data collection activities and introduce our local site coordinators. The letters will describe the study; explain that we have district approval (if applicable); and state the number of study children identified in the school. The letters also will include a description of the study activities and the incentives being offered to teachers for their completion of the survey instruments as well as our assurance of privacy procedures to protect the identity of children and teachers. For schools in systems that approved the study, we will enclose a copy of the signed approval form. For schools where approval remains pending, site coordinators will contact the schools by telephone and/or in person as needed to try to obtain verbal approval directly from the principal and describe the limited burden on each teacher (less than 1 hour in most cases). Then just prior to data collection, our local study staff will contact the schools to confirm that the study children still attend the same classes and make arrangements for the data collection. For children for whom the fall parent tracking update was not completed and/or the current school is not known, the site coordinator and interviewers/assessors will conduct tracing activities to identify and locate the study child's current school setting.

B.4 Tests of Procedures to Minimize Burden

Some instruments proposed for the study were used in the Head Start Impact Study (i.e., Letter-Word Identification, Passage Comprehension, Applied Problems and Calculation from the Woodcock-Johnson Tests of Achievement). By using different starting points, these instruments can be used from preschool age children to adults. The Reading Assessment and the Self-Description Questionnaire (SDQ) from the ECLS-K (Early Childhood Longitudinal Study-Kindergarten Cohort) have extensive psychometric documentation and information on its use with third grade children. The child assessment was pretested with 7 children to verify the time burden on children and the appropriateness of the instruments. The parent interview, teacher survey, and TCRs are similar to those used in the Head Start Impact Study and have been used through four cycles of data collection. The principal survey is similar to the teacher survey, is designed to allow for simple responses and thus minimizing the burden.

B.5 Identity of Individuals Consulted on Statistical Aspects of Design and Identity of Contractors

The contractors for this project are Westat (Contract 232-02-0087) and its subcontractors, Chesapeake Research Associates, Abt Associates, Urban Institute, and the American Institutes for Research.

Project Staff

Michael Puma, Principal Research Associate, Chesapeake Research Associates, (410) 897-4968

Steve Bell, Senior Research Associate, Abt Associates, (301) 634-1721

Gary Shapiro, Senior Statistician, Westat, (301) 517-8028

Pamela Broene, Senior Statistician, Westat, (301) 294-3817

Frank Jenkins, Senior Statistician, Westat, (301) 279-4502

Bibliography

- Abbott-Shim, M., Lambert, R., and McCarty, F. (2003). A comparison of school readiness outcomes for children randomly assigned to a Head Start program and the program's wait list. *Journal of Education for Students Placed at Risk*, 8, 191-214.
- Barnett, W.S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *Future of Children*, 5, 25-50.
- Bickel, R., and Spatig, L. (1999). Early achievement gains and poverty-linked social distress: The case of post-Head Start transition. *Journal of Social Distress & the Homeless*, 8, 241-254.
- Bloom, H.S., (Ed). (2005). *Learning more from social experiments: Evolving analytic approaches*. New York: Russell Sage Foundation.
- Burchinal, M., Roberts, J.E., Zeisel, S.A., Hennon, E.A., and Hooper, S. (2006). Social risk and protective child, parenting, and child care factors in early elementary school years. *Parenting: Science and Practice*, 6, 79-113.
- Butcher, J.N, and Han, K. (1996). Methods of establishing cross-cultural equivalence. In J.N. Butcher (Ed), *International adaptations of the MMPI-2: Research and clinical applications*, (pp. 44-63). Minneapolis, MN: University of Minnesota Press.
- Campbell, F.A., and Ramey, C.T. (1995). Cognitive and school outcomes for high-risk African-American students at middle adolescence: Positive effects of early intervention. *American Educational Research Journal*, 32, 743-772.
- Cauthen, N.K., Knitzer, J., and Ripple, C.H. (2000). *Map and track: State initiatives for young children and families*, National Center for Children in Poverty, Columbia University, NY.
- Curran, P. J. (2000). A latent curve framework for the study of developmental trajectories in adolescent substance use. In J.S. Rose, L. Chassin, C.C. Presson, and S. J. Sherman (Eds.). *Multivariate applications in substance use research: New methods for new questions* (pp. 1-42). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Currivan, Douglas B., Nyman, Amy L., Turner, Charles F., and Biener, Lois. (2004). Does telephone audio computer-assisted self-interviewing improve the accuracy of prevalence estimates of youth smoking? Evidence from the UMass Tobacco Study. *Public Opinion Quarterly*, 68(4):542-564.
- Darlington, R.B., Royce, J.M., Snipper, A.S., Murray, H.W., and Lazar, I. (1980). Preschool programs and later school competence of children from low-income families. *Science*, 208, 202-204.
- Des Jarlais, D. C., Paone, D., Milliken, J., Turner, C. F., Miller, H., Gribble, J., Shi, Q., Hagan, H., Friedman, S. R. (1999). Audio-computer interviewing to measure risk behavior for HIV among injecting drug users: A quasi-randomized trial. *Lancet*; 353: 1657-1661.

- Dickinson, D.K., McCabe, A., Clark-Chiarelli, N., and Wolf, A. (2004). Cross-language transfer of phonological awareness in low-income Spanish and English bilingual preschool children. *Applied Psycholinguistics*, 25(03), 323-347.
- Garces, E., Thomas, D., and Curie, J. (2002). Longer-term effects of Head Start. *The American Economic Review*, 92, 999-1012.
- Gray, S.W., and Klaus, R.A. (1970). The Early Training Project: A seventh year report. *Child Development*, 41, 909-924.
- Hubbs-Tait, L., Culp, A.M., Huey, E., Culp, R., Starost, H. and Hare, C. (2002). Relation of maternal cognitive stimulation, emotional support, and intrusive behavior during Head Start to children's kindergarten cognitive abilities. *Child Development*, 73, 110-131.
- Jo, B. and Muthen, B.O. (2003). Longitudinal studies with intervention and noncompliance: Estimation of causal effects in growth mixture modeling. In . Reise, S. and Duan, N. (Eds.), *Multilevel modeling: Methodological advances, issues, and applications*. Mahwah, NJ, Lawrence Erlbaum Associates.
- Kreisman, M.B. (2003). Evaluating academic outcomes of head start: An application of general growth mixture modeling. *Early Childhood Research Quarterly*, 18, 238-254.
- La Paro, K.M., and Pianta, R.C. (2000). Predicting children's competence in the early school years: A meta-analytic review. *Review of Educational Research*, 70, 443-484.
- Lee, V.E., Brooks-Gunn, J.; Schnur, E., and Liaw, F. (1990). Are Head Start effects sustained? A longitudinal follow-up comparison of disadvantaged children attending Head Start, no preschool, and other preschool programs. *Child Development*, 61, 495-507.
- Metzger DS, Koblin BA, Turner C, et al. Randomized controlled trial of audio computer-assisted self-interviewing: utility and acceptability in longitudinal studies. HIVNET Vaccine Preparedness Study Protocol Team. *American Journal of Epidemiology*, (152) 2, pages 99-106.
- Miles, S.B., and Stipek, D. (2006). Contemporaneous and longitudinal associations between social behavior and literacy achievement in a sample of low-income elementary school children. *Child Development*, 77, 103-117.
- Miller, L.B. and Bizzell, R.P. (1984). Long-term effects of four preschool programs: Ninth- and tenth-grade results. *Child Development*, 55, 1570-1587.
- National Institute of Child Health and Human Development, Early Child Care Research Network. (2005). A day in third grade: A large-scale study of classroom quality and teacher and student behavior. *The Elementary School Journal*, 105.
- National Research Council (1996) *Beyond the blueprint: Directions for research on Head Start's families*. National Academic Press.
- Nelson, G., Westhues, A., and MacLeod, J. (2003). A meta-analysis of longitudinal research on preschool prevention programs for children. *Prevention & Treatment*, 6, no pagination specified.

Peisner-Feinberg, E.S., Burchinal, M.R., Clifford, R.M., Culkin, M.L., Howes, C., Kagan, S.L., and Yazejian, N. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second-grade. *Child Development*, 72, 1534-1553.

Piaget, J. (1983). Piaget's theory. In P. Mussen (ed.). *Handbook of child psychology*. 4th edition. Vol. 1. New York: Wiley.

Pogrow, S. (2000, April 19). Beyond the "good start" mentality. *Education Week*.

Ramey, C.T., and Ramey, S.L. (2004). Early learning and school readiness: Can early intervention make a difference? *Merrill-Palmer Quarterly*, 50, 471-491.

Raudenbush, S., and Bryk, A. (2002). *Hierarchical linear models: Applications and analysis methods, 2nd Edition*. Thousand Oaks, CA: Sage Publications.

Raudenbush, S.W. (2003). Designing field trials of educational innovations. Paper presented at the national invitational conference, Conceptualizing Scale-Up: Multidisciplinary Perspectives, Nov. 3, 2003, Washington, DC.

Raudenbush, S.W., Bryk, A.S., Cheong, Y.F., and Congdon, R. (2005). *HLM 6: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.

Reynolds, A. J.; Ou, S., and Topitzes, J. W. (2004). Paths of effects of early childhood intervention on educational attainment and delinquency: A confirmatory analysis of the Chicago Child-Parent Centers. *Child Development*, 75, 1299-1328.

Reynolds, A. J. (1999). *The Chicago longitudinal study: A study of children in the Chicago Public Schools*.

Rubin, D.B. (2004). Direct and indirect causal effects via potential outcomes. *Scandinavian Journal of Statistics*, (31), pages 161-170.

Schulman, K., Blank, H., and Ewen, D. (1999). *Seeds of success: State prekindergarten initiatives 1998-1999*. Washington, DC: Children's Defense Fund.

Teske, P., Wolf, P., and Hill, P. (2006). Poor parents can be good choosers. *Education Week*, 25(32), 44.

Turner, C. F., Ku, L. , Rogers, S. M., Lindberg, L. D., Pleck, J. H., Sonenstein, F. L. (1998). Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science*, (280).5365, pages 867 – 873.

U.S. Department of Health and Human Services. Advisory Committee on Head Start Research and Evaluation. (1999). *Evaluating Head Start: A recommended framework for studying the impact of the Head Start program. Advisory committee on Head Start research and evaluation*. Washington, DC: U.S. Government Printing Office.

U.S. Department of Health and Human Services. (2003).. *The Head Start Impact Study: Research design and preliminary analysis plan*. Report prepared by Westat and its subcontractors under contract 282-00-0022 .

U.S. Department of Health and Human Services. (2005). *Technical note #4: Head Start Impact Study: Getting the research questions right*. Report prepared by Westat and its subcontractors under contract 282-00-0022 .

U.S. General Accounting Office. (1997). *Head Start: Research provides little information on impact of current program*. Washington, DC: U.S. Government Printing Office.

U.S. General Accounting Office. (1998). *Head Start: Challenges in monitoring program quality and demonstrating results*. Washington, DC: U.S. Government Printing Office.

Wolf, M., and Stein, A. (1966). *Study I: Six months later, a comparison of children who had Head Start, summer 1965, with their classmates in kindergarten (A case study of kindergartens in four public elementary schools, New York City.)* Washington, DC: Research and Evaluation Office, Project Head Start, Office of Economic Opportunity.

Zill, N., Resnick, G., Kim, K., O'Donnell, K., Sorongon, A., McKey, R., Pai-Samant, S., Clark, C., O'Brien, R., and D'Elio, M. (2003). *Head Start FACES 2000: A whole-child perspective on program performance. Report prepared for the Administration for Children and Families, U.S. Department of Health and Human Services (DHHS) under contract HHS-105-96-1912, Head Start Quality Research Consortium's Performance Measure Center.*

Appendix A

Third Grade Follow-Up Study Instruments

INCLUDED IN A SEPARATE FILE

Appendix B

Federal Register Announcement

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Administration for Children and Families

Proposed Information Collection Activity; Comment Request
Proposed Projects

Title: Third Grade Follow-up to the Head Start Impact Study.

OMB No.: 0970-0229.

Description: The Administration for Children and Families (ACF) within the Department of Health and Human Services (HHS) is requesting comments on plans to implement a third grade follow-up to the Head Start Impact Study. This study will collect information for determining, on a national basis, how Head Start affects outcomes in the third grade for children who participated in the program as compared to children not enrolled in Head Start and to determine under which conditions Head Start works best and for which children.

The Head Start Impact Study was a longitudinal study that involved approximately 5,000 first-time-enrolled three- and four-year-old pre school children across 84 nationally representative grantee/delegate agencies (in communities where there were more eligible children and families than can be served by the program). The participating children were randomly assigned to either a Head Start group (that could enroll in Head Start services) or a control group (that could not enroll in Head Start services but could enroll in other available services selected by their parents). Data collection for the study began in the fall of 2002 and extended through spring 2006.

It is the intention of the Administration for Children and Families to examine outcomes for this sample of children and families during the spring of the children's third grade year. Data will be collected in the spring of 2007 (for the four-year-old cohort) and the spring of 2008 (for the three-year-old cohort). The domains for development to be assessed include demographic characteristics of the children and families, as well as children's cognitive development, school achievement and adjustment, socio-emotional functioning, health and access to health care, and relationships with peers. Information will also be collected on parents' involvement in educational activities, mental health and well-being, and monitoring and other parenting practices, and information related to the characteristics and quality of the schools and classrooms that children attend.

Respondents: Individuals or households and school districts.

Annual Burden Estimates

Number of Average

respondent	Responses per respondent	Respondents and activities per response	burden hours	Total burden hours	Number of respondents
1	1	Child Interview and Assessment.....	1	4,600	4,600
1	1	Parent Interview.....	1	4,600	4,600
1	.33	Teacher Survey.....	.33	1,533	4,600
1	.25	School Administrator Survey.....	.25	575	2,300

Total Annual Burden					
Estimates.....					
11,308					

[[Page 46917]]

In compliance with the requirements of Section 3506(c)(2)(A) of the Paperwork Reduction Act of 1995, the Administration for Children and Families is soliciting public comment on the specific aspects of the information collection described above. Copies of this proposed collection of information can be obtained and comments may be forwarded by writing to the Administration for Children and Families, Office of Information Services, 370 L'Enfant Promenade, SW., Washington, DC 20447, Attn: ACF Reports Clearance Officer. All requests should be identified by the title of the information collection. E-mail: infocollection@acf.hhs.gov.

The Department specifically requests comments on (a) whether the proposed collection of information is necessary for the proper performance of the functions of the agency, including whether the information shall have practical utility; (b) the accuracy of the agency's estimate of the burden of the proposed collection of information; (c) the quality, utility, and clarity of the information to be collected; and (d) ways to minimize the burden of the collection of information on respondents, including through the use of automated collection techniques or other forms of information technology. Consideration will be given to comments and suggestions submitted within 60 days of this publications.

Dated: August 9, 2006.
 Robert Sargis,
 Reports Clearance Officer.
 [FR Doc. 06-6924 Filed 8-14-06; 8:45 am]
 BILLING CODE 4184-01-M

Appendix C

Public Comments and Response

ACF Response to Comments

The National Head Start Association (NHSA) makes a number of important points in its comment that will be addressed here:

- 1) The importance of assessing the quality of schools Head Start children attend using observational techniques;
- 2) The importance of examining how various environmental risk factors contribute to the development of Head Start children and families;
- 3) The importance of examining grade retention and achievement test data.

The Administration for Children and Families agrees that the quality of the school that children attend is of great importance to the Head Start Impact Study. In requesting proposals for this work, it requested that bidders provide details and costs for conducting observations of classroom quality within all, or a sample, of classrooms attended by the children in the study. Unfortunately, the proposals offered by the bidders indicated that this effort would be extraordinarily expensive – increasing the total cost of the contract by 30% -- and yet would have resulted in Minimum Detectable Effects that were too high for this study. Thus, the effort would add tremendously to the overall cost of the study, but would be unlikely to identify effects of the size that have been identified in the first report for this study. As a result, ACF has decided that the benefits of collecting observational data do not outweigh their costs. Instead, this study will use surveys with teachers and principals to collect information on school and classroom quality. While not ideal, this technique will provide useful information for analyzing the question of how school quality relates to Head Start program impacts.

The second recommendation from the National Head Start Association was for an analysis of how various risk factors contribute to the development of Head Start children and families. This study will examine this issue in a few ways: first, to the degree that children in the program and control groups differ on any characteristics such as those mentioned by NHSA, they will be included as covariates in the impacts analyses; second, this study will examine the degree to which impacts on children and families differ based on the presence of such risk factors in their environments. However, ACF does not intend to conduct a basic analysis of the relationship between environmental and family characteristics and children's outcomes within this study. Although this is a very important topic for consideration, it does not directly relate to the research question of this study, which emphasize the impact of Head Start. Since the study used a randomized control trial, there should be minimal need to consider the role of environmental characteristics in children's development, except as they indicate key subgroups who might experience differential impacts. ACF has a number of other studies – the FACES study, of greatest relevance – that continue to examine questions related to how characteristics of Head Start children's families and communities contribute to their development.

Finally, ACF agrees with the NHSA that it is important to examine both grade retention and children's achievement test scores for their potential impact of Head Start. This study will examine both, through teacher reports of children's grade retention and through standardized assessments of children's math and reading skills.

Appendix D

Advisory Committee Members

Members of the Advisory Committee on Head Start Research and Evaluation*

October 19, 1999

(Prior to the Head Start Impact Study award)

Martha S. Abbott-Shim, Georgia State University
J. Lawrence Aber, Columbia University
Rosalinda B. Barrera, University of Illinois at Urbana-Champaign
C. Hendricks Brown, University of South Florida
Donna M. Bryant, University of North Carolina-Chapel Hill
Margaret R. Burchinal, University of North Carolina-Chapel Hill
Thomas D. Cook, Northwestern University
Gayle Cunningham, Jefferson County Committee for Economic Opportunity Community Action Agency, Birmingham, Alabama
Greg J. Duncan, Northwestern University
Harriet C. Ganson, U.S. General Accounting Office
Olivia Golden, U.S. Department of Health and Human Services, Administration for Children and Families
Richard F. Gonzales, Administration for Children's Services-Head Start, New York City
Sarah M. Greene, National Head Start Association
Wade F. Horn, National Fatherhood Initiative
Jacqueline Jones, Educational Testing Service
Joan Lombardi, Yale University
John Love, Mathematica Policy Research
Patricia Montoya, U.S. Department of Health and Human Services, Administration on Children, Youth, and Families
Suzanne M. Randolph, University of Maryland
Craig T. Ramey, University of Alabama at Birmingham
Peter H. Rossi, University of Massachusetts at Amherst (retired)
Lawrence J. Schweinhart, High/Scope Educational Research Foundation
Ann Segal, U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation
Robert G. St. Pierre, Abt Associates, Inc.
Helen H. Taylor, U.S. Department of Health and Human Services, Head Start Bureau
Grover J. Whitehurst, State University of New York at Stony Brook
Diane J. Willis, University of Oklahoma
Hirokazu Yoshikawa, New York University
Edward F. Zigler, Yale University
Nicholas Zill, Westat, Inc.

*Affiliation is reported at the time of the Committee appointment.

Members of the Advisory Committee on Head Start Research & Evaluation*
January 12, 2001

Martha Abbott-Shim, Georgia State University
Lawrence Aber, Columbia University
Rosalinda Barrera, University of Illinois--Urbana-Champaign
Barbara Broman, U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation
C. Hendricks Brown, University of South Florida
Donna M. Bryant, University of North Carolina-Chapel Hill
Margaret R. Burchinal, University of North Carolina-Chapel Hill
Thomas D. Cook, Northwestern University
Gayle Cunningham, Jefferson County Committee for Economic Opportunity, Birmingham, Alabama
Greg J. Duncan, Northwestern University
Sarah Greene, National Head Start Association
Olivia Golden, U.S. Department of Health and Human Services, Administration for Children and Families
James Harrell, U.S. Department of Health and Human Services, Administration for Children and Families
Robinson Hollister, Swarthmore
Wade Horn, National Fatherhood Initiative
Andrew Kennedy, Los Angeles County Office of Education
Doug Klafehn, U.S. Department of Health and Human Services, Head Start Bureau
Joan Lombardi, Yale University
John Love, Mathematica Policy Research
Timothy Nolan, Waukesha (Wisconsin) County Project Head Start, Inc.
Suzanne Randolph, University of Maryland
Peter Rossi, University of Massachusetts at Amherst (retired)
Lawrence Schweinhart, High/Scope Educational Research Foundation
Robert St. Pierre, Abt Associates, Inc.
Grover J. Whitehurst, State University of New York at Stony Brook
Karen Whiten, U.S. General Accounting Office
Diane J. Willis, University of Oklahoma
Edward F. Zigler, Yale University

*Affiliation is reported at the time of the Committee appointment.

Members of the Advisory Committee on Head Start Research & Evaluation*
May, 2003

Mark, Appelbaum, University of California--San Diego
Margaret Burchinal, University of North Carolina--Chapel Hill
Nell Carvell, Southern Methodist University
Gayle Cunningham, Jefferson County Committee for Economic Opportunity, Birmingham,
Alabama
Sarah Greene, National Head Start Association
Ron Haskins, The Brookings Institute
Mary Louise Hemmeter, University of Illinois at Urbana--Champaign
Windy M. Hill, U.S. Department of Health and Human Services, Administration for Children,
Youth and Families
Wade Horn, U.S. Department of Health and Human Services, Administration for Children and
Families
Nicholas Ialongo, Johns Hopkins University
Marci Kanstoroom, Thomas B. Fordham Foundation
Kristin Moore, Child Trends
Joan E. Ohl, U.S. Department of Health and Human Services, Administration for Children, Youth
and Families
David Olds, University of Colorado
Craig Ramey, Georgetown University
Howard Rolston, U.S. Department of Health and Human Services, Administration for Children
and Families
William Rosenberg, Drexel University
Ann Segal, Annie Casey Foundation
George Smith, Management Planning Institute, Inc.
Dorothy Strickland, Rutgers University
William Douglas Tynan, Al duPont Hospital for Children
Betty Ward-Zukerman, Education, Workforce, and Income Security Team
Don Winstead, U.S. Department of Health and Human Services, Office of the Assistant Secretary
for Planning and Evaluation
Edward Zigler, Yale University

*Affiliation is reported at the time of the Committee appointment.

Appendix E

Head Start Impact Study Sampling Plan

B.2 Description of Sampling and Information Collection Procedures

Sampling Procedures

Two principles guided the development of this sampling strategy:

- **National Representation**—the sample needed to support our ability to extrapolate estimated program impacts to the population of all Head Start programs and children.
- **Creation of a Randomized Comparison Group**—sampled Head Start programs had to be able to provide a sufficient number of additional applicants to allow the use of an experimental design, i.e., the selected programs had to have more applicants than could be served.

The process used to select and recruit the study sample, under these two overarching principles, is described below and summarized in Exhibit 6.

Legislative Mandate and Advisory Committee Guidance. The legislative mandate required that the Head Start Impact Study provide:

“...a national analysis of the impact of Head Start” based on the selection of Head Start grantees/delegate agencies that “...operate in the 50 states, the Commonwealth of Puerto Rico, or the District of Columbia and that do not specifically target special populations.”

Furthermore, the Advisory Committee recommended that the sample of Head Start grantees/delegate agencies should reflect variation in a variety of characteristics including,

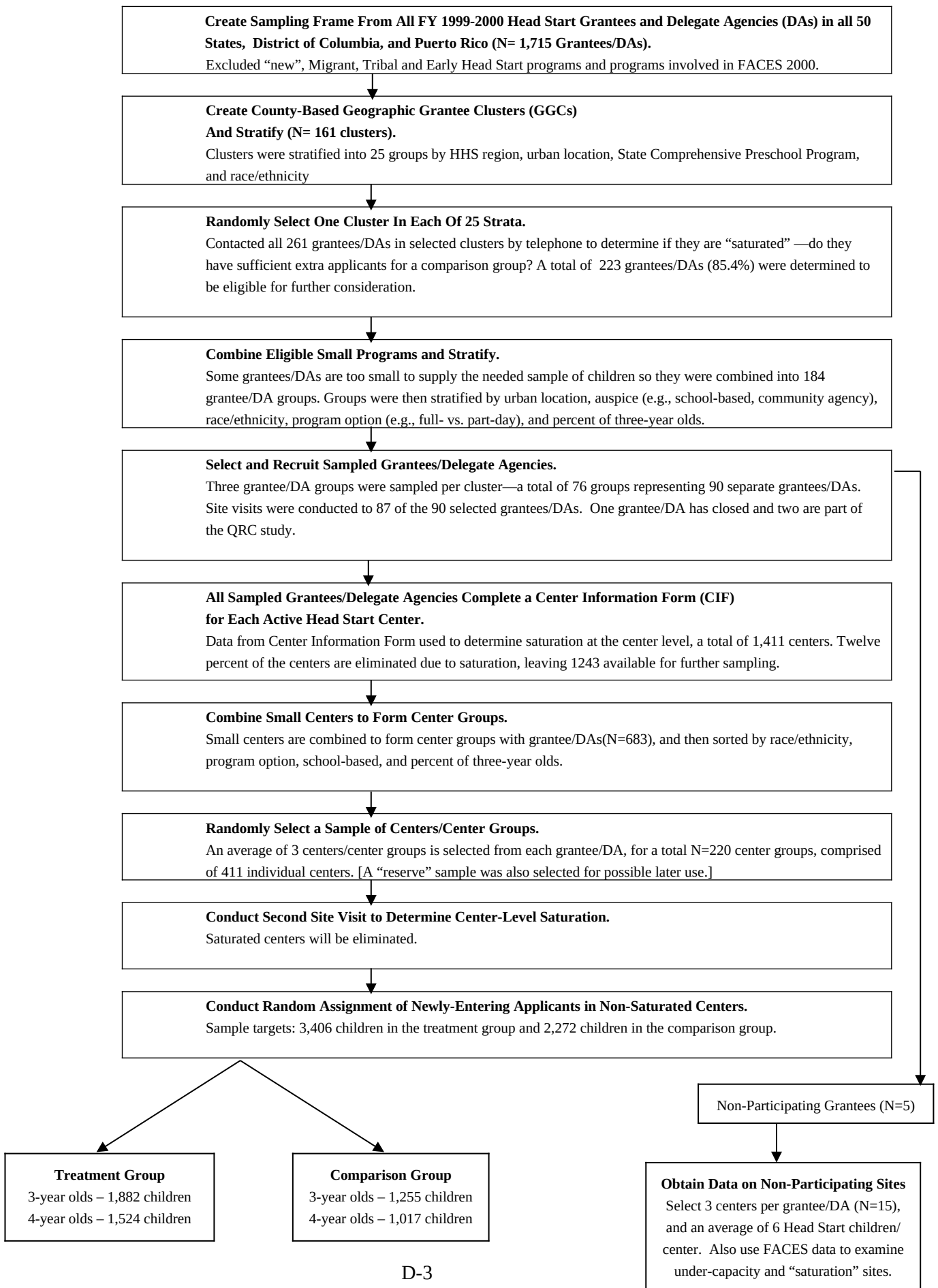
“...region of the country, race/ethnicity/language status, urban/rural, and depth of poverty in communities,” and “...design of program as a one-year or two-year experience for children; program options (e.g., center-based, home-based, part-day, full-day); auspice (e.g., Community Action Agency, public school, non-profit organization);

community-level resources; alternative childcare options for low-income children; and, the nature of the childcare market and the labor market in the community studied.”

The sampling plan, therefore, began with the inclusion of all operating Head Start grantees/delegate agencies listed in the 1999 and 2000 Head Start Program Information Report (PIR), excluding two types of programs that target “special populations”—those serving migrant children, and those operated by Tribal Organizations—and those programs that are “extremely new to the program” because they may not represent stable Head Start operations. Children enrolled in Early Head Start (i.e., those younger than three years of age) were also excluded from the study sample because they are assured entry into regular Head Start (and cannot, therefore, be randomly assigned).

Geographically Cluster Grantee/Delegate Agencies. Using the resulting list, grantees/delegate agencies were clustered by county on the basis of their primary business address. Every county in the US with at least one operating grantee was included in the sampling frame with a minimum of at least eight (8) grantees/delegate agencies included per cluster. Very

Exhibit 6: Overall Plan for Sample Selection



small grantees/delegate agencies, with fewer than 90 total 3- and 4-year old Head Start children, were combined with another grantee/delegate agency in the same county if possible or with one in an adjacent county. A total of 160 clusters were formed.

Stratify the Sample to Ensure National Program Representation. Next, the clusters were combined into 25 strata, each stratum having approximately the same number of three- and four-year old Head Start children. The following stratification variables were used:

- **Region.** Defined as five regions⁵: (1) **Northeast**, Head Start Regions 1, 2, and 3; (2) **South**, Head Start Regions 4 and 6; (3) **North Central**, Head Start Region 5; (4) **Plains**, Head Start Regions 7 and 8; and (5) **West**, Head Start Regions 9 and 10.
- **Urban Location.** Three categories were defined⁶: (1) a county **containing a central city of a Metropolitan Statistical Area (MSA)** with 1 million or more persons; (2) a county in an MSA not included in the first category (i.e., a suburban county or any county in a small MSA); and, (3) all other areas of the country (i.e., areas not in an MSA, predominantly small towns and rural).
- **State Comprehensive Programs for Low-Income Preschool Children.** Three groups: (1) States with comprehensive state-funded pre-kindergarten programs **that are similar to Head Start**; (2) States with state-funded pre-kindergarten programs that have **some comprehensive program components**; and, (3) States meeting neither of the previous two requirements.
- **Race/Ethnicity.** Three categories: (1) **High concentration of Hispanic Head Start children**, the percentage of Hispanic children served by the grantees/delegate agencies in the cluster is at, or above, 40 percent; (2) **High concentration of African American children (but not of Hispanic children)**, the percentage of non-Hispanic African American children served by the grantees/delegate agencies in the cluster is at, or above, 40 percent and the percentage of Hispanic children below 40 percent; and, (3) **Other**, all other clusters not included in the preceding categories.

⁵ **Northeast:** Maine, New Hampshire, Vermont, Massachusetts, Connecticut, Rhode Island, New York, New Jersey, Puerto Rico, Pennsylvania, Delaware, Maryland, West Virginia, Virginia, and the District of Columbia, a total of 15 states. **South:** North Carolina, South Carolina, Georgia, Florida, Mississippi, Tennessee, Kentucky, Alabama, Louisiana, Oklahoma, Texas, New Mexico, and Arkansas, a total of 13 states. **North Central:** Ohio, Indiana, Illinois, Michigan, Wisconsin, and Minnesota, a total of 6 states. **Plains:** Nebraska, Iowa, Missouri, Kansas, North Dakota, South Dakota, Montana, Wyoming, Colorado, and Utah, a total of 10 states. **West:** California, Arizona, Nevada, Idaho, Washington, Oregon, Alaska, and Hawaii, a total of 8 states.

⁶ Defined using Census Bureau designations for Metropolitan Statistical Areas (MSA's) and urbanicity (i.e., Beale Codes).

These variables were used to create 25 strata each of which included about the same number of Head Start children.

Select Sample of Geographic Grantee Clusters. Once the strata were formed one cluster from each stratum was selected with probabilities proportional to size, i.e., clusters with larger numbers of Head Start children had a higher probability of being selected into the sample. This plan ensured that each Head Start grantee/delegate agency and participating child has a known probability of selection into the study sample, and—at later points in the sampling process—that the probabilities will be approximately the same for each child. The decision to sample a total of 25 clusters was based on a tradeoff between (1) the need to control the cost and quality of data collection and our ability to monitor random assignment (argues for fewer clusters), and (2) concerns about the magnitude of the confidence intervals around the estimated program impacts (argues for more clusters).

The 25 clusters included 355 grantee/delegate agencies. At this point, eight grantees/delegate agencies that were involved in the FACES 2000 study were dropped.

Identify Grantees/Delegate Agencies Eligible For The Study. In each of the 25 clusters, an attempt was made to contact all grantees/delegate agencies by telephone, except in three very large clusters—Los Angeles, Chicago, and Brooklyn—where twelve programs (or groups of programs) were randomly subsampled in each cluster prior to screening to reduce the amount of telephone calls that would have to be made. Eighty six additional grantees/delegate agencies were eliminated. In all, an attempt was made to screen a total of 261 programs.

Information gathered during the telephone calls was used to determine if the grantee/delegate agency was still actively serving three- and/or four-year old children, and “not saturated,” i.e., if it had enough applications from **newly-entering** children in one or more centers (possibly only in one service option) to fill **all** federally-funded Head Start slots **and** have enough applicants left over for a comparison group.

Of the 261 programs, a total of 223 (85%) were either determined to be eligible, or their eligibility could not be established because not enough information was provided (this included a total of 28 programs that remained in the pool for possible selection).

Collapsing, Stratifying, and Selecting Grantees/Delegates Within Sampled Clusters. As in an earlier step, small grantees/delegate agencies were combined to avoid a sample shortfall of children (and to avoid the added variability and response burden that would arise if all children in a small program had to be taken with certainty). This procedure reduced the total sampling frame from 223 to 184 stand-alone grantees/delegate agencies or combined groups (groups are treated as a single unit for sampling purposes).

The 184 grantee/delegate agency groups were then stratified to ensure representation of important groups of programs and children. The variables used for this purpose (from highest to least importance) were:

- Urban Location (Central City, Other Urban, Rural/Small Town)
- Auspices (School based vs. all other agency types)
- Percent Hispanic enrollment (> 40% vs. ≤ 40%)
- Percent Black enrollment (> 40% vs. ≤ 40%)
- Program Option (Part day only, Full day only, Both)
- New 3-Year Old Enrollment as a percent of total new 3- and 4-year old enrollment

Within each sampled cluster, grantees/delegate agency groups were sorted by the stratification variables, and an initial sample of 3 programs per cluster was selected with probability proportional to the total number of newly entering three- and four-year olds. A total of 90 individual programs were selected for the main sample, representing a total of 76 grantee/delegate agency groups. (Exhibit 7 shows the extent to which the sample is representative of the frame across the stratification variables).

Three of the 90 programs were eliminated prior to beginning recruitment. One program closed and two others were part of another current Head Start study, the Quality Research Center Study. Each of these programs was part of a grantee/delegate agency group, so all 76 groups remain with a total of 87 programs.

Exhibit 7:

Distribution of Sample vs. Frame vs. PIR, Weighted by Total Age 3 to 5 Enrollment

Variable	Percent for Sample (n=90)⁷	Percent for Frame (N=223)⁸	Percent for PIR (N=1715)
Urban Location ⁹			
Central City	49	46	40
Other Urban	30	31	35
Rural/Small Town	21	23	25
Head Start Region			
Northeast	25	25	23
South	39	41	34
North Central	22	19	20
Plains	4	3	7
West	10	12	16
Auspices			
School-based	7	10	13
Other	93	90	87
Pre-K Program State Expenditures			
Has comprehensive programs similar to Head Start	24	27	28
Has some comprehensive programs similar to Head Start	21	20	18
Remaining States	55	53	54
Percent Hispanic			
High (> 40%)	28	31	28
Low (≤ 40%)	72	69	72
Percent Black			
High (> 40%)	44	42	41
Low (≤ 40%)	56	58	59
Hispanic Enrollment	32	31	28
Black Enrollment	39	39	37
Other Enrollment	29	30	35
Age 3 Enrollment	35	34	35
Age 4 Enrollment	62	61	60
Age 5 Enrollment	3	5	5

⁷ Also weighted by PSU and program weight to reflect PSU and program probabilities of selection

⁸ Also weighted by PSU weight to reflect PSU probabilities of selection and by inverse of probability of selection for subsampling programs within LA county, Chicago, and Brooklyn prior to screening.

⁹ Using Beale code: 0 = Central City, 1-3 = Other Urban, 4-9 = Rural/Small Town

Initial Recruitment Contacts. The core of the approach in working with grantees/delegate agencies selected for the main study was to establish strong partnerships with the grantees/delegate agencies, actively address potential concerns, and work to mitigate Head Start staff and study participants' concerns regarding issues such as random assignment, to the extent possible. Additionally, our goal was to coordinate program recruitment efforts in close collaboration with regional office staff in order to minimize any potential confusion with the requirements of study participation and normal program compliance or monitoring requirements.

To accomplish these goals, study staff were divided into recruitment teams, with each team responsible for recruiting about ten grantees. The recruitment process involved ongoing telephone and in-person contact with each grantee and their selected centers. On average, three site visits were made to each site at strategic points in the recruitment process. Each grantee and associated center also had a local site coordinator assigned to maintain ongoing personal contact to allay concerns about random assignment, help gather the information necessary to conduct random assignment, and to eventually supervise the data collection.

Once the 87 grantees were selected, letters, followed by telephone calls were made to grantee directors to begin to engage their cooperation. Site visits were conducted in the summer of 2001. All 87 grantees were visited and provided information about the study and its random assignment design. Meetings were held with key personnel at the grantee level and initial participation agreement was sought. Overall there was acceptance of the study and willingness on the part of all but two grantees to continue negotiating the conditions of participation. Various issues were raised, but the grantees were very willing to continue the dialogue to address the issues. Our approach was to consider ways in which the procedures can be adjusted to meet individual site needs while at the same time maintain the integrity of a random assignment design.

A key to successful site recruitment is to ensure program administrators that the use of random assignment will not impose too many burdens on potential participant families, nor generate dissatisfaction in this vital client population. The largest step toward this goal was the decision to conduct random assignment only in Head Start grantees/delegate agencies operating at or near full capacity and where there was an additional pool of unserved families in the community known to be interested in services. This ensured that the evaluation randomly assigned children to the control group only in communities where Head Start grantees/delegate agencies currently cannot serve all eligible applicants due to limitations in the number of funded slots.

Selecting Centers Within Grantees/Delegate Agencies. After the first site visit, all of the recruited grantees/delegate agencies were asked to complete a Center Information Form (CIF) in late 2001 for each of their currently operating centers, as well as for any centers that were expected to be in operation as of Fall 2002. This information provided center-level verification of recruitment and enrollment data comparable to PIR data and was used to determine if a particular center could supply the “extra” applicants needed for the comparison group. Centers were initially eliminated from further consideration using the grantee-reported information if they met **all three** of the following conditions:

1. **Overall Under-enrollment:** Center is under-enrolled by more than 10%, i.e., $[(\text{Total slots}) - (\text{Total enrollment})] > 10$ percent of Total slots.
2. **No Waiting List:** There are no income-eligible children on the center’s waiting list.
3. **No fully-enrolled options:** None of the center’s available service options are fully enrolled.

This was a process that tried to eliminate centers that were clearly saturated without eliminating too many centers on the basis of what is likely to be uncertain information. 1,411 center information forms were submitted and this step eliminated about 168 centers, (approximately 12 percent of all centers). Regional offices were also consulted about the enrollment numbers of centers. A list of centers, along with any saturation concerns were sent to regional office staff for their verification.

Forming Center Groups. The sample design called for sampling 16 treatment and 11 control children in their first year of Head Start per center. However, only about 48% of the centers mailing in CIFs reported a first year enrollment of at least 27 children. As with grantees/DAs, “small” centers (i.e., those with relatively few newly-entering children) were combined into “center groups,”¹⁰ each with a combined reported first year enrollment of at least 27. The likelihood of each center providing the required sample sizes of treatment and control children was taken into consideration in forming center groups. If a center was underenrolled relative to the number of funded slots allocated to it, or if it reported having no waiting list, it was more likely to be combined with another center, even if its reported first year enrollment exceeded 27. Geographical proximity was also a consideration in grouping centers. Each center was ranked from 1 to 8, with 1 indicating the center was very likely to meet the study requirements and 8 indicating very unlikely. Centers with a score of 8 were deleted from the frame. Each center group was treated as a single unit for sample selection, and the average target sample size of program and comparison group children was sampled from the combined unit.

¹⁰ To protect ourselves against unexpected problems, those centers that appeared to be “almost saturated” from the previous step were given a reduced estimated number of newly-entering children (this reduces their probability of being selected, and often results in their being combined with other centers thereby expanding the potential pool of study children).

Stratification of Centers. The center groups were implicitly stratified within each program prior to sampling by sorting them. When used in conjunction with systematic sampling, sorting improves the representativeness of the sample. We used implicit rather than explicit stratification because the number of potential strata in each program generally exceeded the target sample size. Note, however, that in many programs the sorting variables were not especially useful because of the extremely high sampling fractions in most center groups.

The variables used for sorting were obtained from the CIF. These variables were, in order of priority, 1) whether the center is school-based (i.e. teachers are hired by the school district:Yes/No), 2) the percent Hispanic enrollment and/or the percent Black enrollment, 3) program option (full-day, part-day, both/home-based/other), and 4) the percent of first year three year old enrollment. When the school-based or program option variables differed for centers comprising a center group, the group was assigned to the category with the largest first year enrollment. In programs where there was little or no variation on a particular variable, the variable wasn't used for sorting. In a few programs, the priority of the sorting variables was altered to reflect the particular distribution of centers within the program. When the percent Hispanic or percent Black were not the last sort variables, they were categorized into High/Low based on a cutoff which could differ for each program, depending on the distribution of race/ethnicity in the program.

Initial Center Selection and Screening. A larger than needed sample of centers (or center groups) was randomly selected from each grantee/DA, up to about 6 centers/groups per program. Centers/groups were selected with probabilities proportional to each center's estimated enrollment of newly-entering three- and four-year old children (i.e., larger centers/groups had a higher chance of being selected). We then randomly selected an average of 3 centers/groups per grantee/DA to serve as a "main" sample with the remainder serving as a "reserve" to be used if we can't obtain our desired sample of children from particular types of centers, or for particular types of children, **across the entire study**, not just from within any given grantee or cluster. Very small programs were given a minimum allocation of one center to ensure they could participate in the study, since they had already been recruited. In a number of programs the target sample size exceeded the number of center groups available for sampling, due to extensive collapsing in the program. In these programs we sampled all center groups, and there was no reserve sample.

When the target sample size did not exhaust the number of center groups available, we assigned a reserve sample size of two center groups if the target was less than five, and three if the target was five or more. In a few programs there was only one center group available for the reserve sample.

The sample was drawn in two stages in order to select both a main and a reserve sample. At the first stage, the center groups were first sorted within each program, then sampled with probability proportional to the first year enrollment for the center group. The first stage sample size was calculated to be large enough to accommodate both the target and reserve sample sizes. At the second stage, the first stage sample was sorted in the order of selection and an equal probability sample of center groups was selected. The sample size at the second stage was equal to the target. Center groups selected in the initial sample but not at the second stage became the reserve sample.

In programs where all first stage center groups were sampled with certainty (i.e. a census) but the target sample size did not exceed the number of center groups available, a reserve sample was possible and a second stage sample was selected as described above. In other programs the first stage sample contained a mixture of certainty and noncertainty selections due to a large measure of size for some center groups. In these programs the certainty selections at the first stage were retained for the main sample, and the remaining number needed to complete the target sample size were sampled at the second stage. The certainties from the first stage and center groups subsampled at the second stage are the main sample, and the remaining center groups not selected at the second stage are the reserve sample.

Two centers were given a zero measure of size to avoid sampling them at the request of the ACYF, to avoid conflict with another Head Start study (QRC). An adjustment to the center base weights will be done to compensate for their exclusion from the frame. Center sampling was completed in 85 of the 90 main sample Head Start programs.

Final Center Selection. A second round of discussions and site visits conducted in January-April, 2002 focused on the centers selected for the “main sample” to determine if they were able to provide the desired sample of treatment and comparison group children (i.e., they are not “saturated”). This involved verifying enrollment information for each center/group. The results of this screening were well documented before we decided to eliminate a particular center from further consideration. It is also important to note that for the small number of centers that have multiple service options (e.g., full- and part-time), we sought to conduct random assignment for any un-saturated service option(s). In a similar manner, we conducted random assignment for only one age group (e.g., only three-year olds) if the other

was found to be saturated due, for example, to competition for four-year olds among preschool programs in the community.

Select Appropriately-Sized Samples of Head Start Children. In the selected Head Start centers, spread across the sample of study grantees/delegate agencies, we proposed to select an initial sample of 3,137 newly entering three-year-old participants and 2,541 newly entering four-year-old participants. As shown in Exhibit 8, we anticipate that a total of 1,882 three-year-olds will be assigned to the treatment group and 1,255 to the comparison group, while a total of 1,524 four-year-olds will be assigned to the treatment group and 1,017 4-year-olds to the comparison group. This exhibit also indicates the anticipated sample sizes for each wave of data collection. To obtain a final sample of 1,667 three-year-olds and 1,667 four-year-olds at the end of the study period—the size needed for adequate statistical precision—we estimate a beginning sample size of 3,137 three-year-olds and 2,541 newly enrolled four-year-olds.

Exhibit 8

Expected Sample Size At Each Wave Of Data Collection¹¹

COHORT 1: Two Year Participants (Three-Year-Olds)

	Participating Grantees/Delegate Agencies		
	Treatment	Comparison	Total
At Random Assignment	1,882	1,255	3,137
Fall 2002 HS	1,694	1,130	2,824
Spring 2003 HS	1,524	1,017	2,541
Fall 2003 HS	1,372	915	2,287
Spring 2004 HS	1,235	823	2,058
Spring 2005 K	1,111	741	1,852
Spring 2006 1 st grade	1,000	667	1,667

COHORT 2: One Year Participants (Four-Year-Olds)

	Participating Grantees/Delegate Agencies		
	Treatment	Comparison	Total
At Random Assignment	1,524	1,017	2,541
Fall 2002 HS	1,372	915	2,287
Spring 2003 HS	1,235	823	2,058
Spring 2004 K	1,111	741	1,852
Spring 2005 1 st grade	1,000	667	1,667

Final Recruitment Steps. The final steps of recruitment included verifying any saturation problems of selected centers, gaining center staff cooperation, developing specific random assignment

¹¹ Includes an assumed 10% attrition rate each year.

procedures, and coming to a final agreement about study procedures in each individual site. Site visits were conducted in January – April 2002 and when needed, again in the early summer of 2002. Partnership agreements were signed with each grantee/delegate agency, outlining agreed upon study procedures. Random assignment procedures were implemented in each site as soon as their enrollment decisions were made. The recruitment process entailed ongoing contact with the sites via personal visits and telephone calls. Each recruitment effort worked through any potential concerns about participating in the study, develop individualized study plans with the grantees, and obtained information on the community context.