Head Start Family and Child Experiences Survey (FACES 2009)

Supporting Statement Part B for OMB Approval

February 27, 2009

B. STATISTICAL METHODS (USED FOR COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS)

B.1. Respondent Universe and Sampling Methods

The sample design for the new cohort of the Head Start Family and Child Experiences Survey (FACES 2009) is essentially the same as the one used for FACES 2006, which was in turn based on the designs of the three prior cohorts. In the interest of continuity, and because this sample design worked well, we saw no reason to change the basic design. However, in FACES 2009 we decided to minimize the burden on parents/guardians who had more than one child selected for the sample—and thus in FACES 2006 had to do two parent interviews at each data collection point, one for each selected child. We now propose to randomly subsample one selected child per parent/guardian. In addition, at the request of ACF, we now plan to follow children who leave Head Start prior to their kindergarten year. In FACES 2006, these "leavers" were considered ineligible for data collection, because they did not have a full year of Head Start prior to kindergarten. Because a growing proportion of Head Start program users leave the program after one year for state-funded pre-kindergarten or other such programs, ACF would like to consider these children as part of the Head Start FACES population and follow them through kindergarten along with the rest of the original sample. Additionally, in FACES 2006 we oversampled 3-year-olds in order to ensure sufficient sample size for analyses. We found in FACES 2006 that there were sufficient 3-year-olds without oversampling, therefore we will not oversample 3-year-olds in FACES 2009.

FACES 2009 will use a stratified multi-stage sample design with four stages: (1) Head Start programs, with programs defined as grantees or delegate agencies providing direct services; (2) centers within programs; (3) classes within centers; and (4) children within classes. We will select a probability sample of programs using the 2007-2008 Head Start Program Information Report (PIR) as the sampling frame, which contains information provided by each program each year. We will exclude from the sampling frame: programs in Puerto Rico and U.S. territories, migrant and American Indian programs, programs not directly providing services to children in the target age group, and programs that are (or will soon be) defunded. We may also exclude programs or centers that are currently participating in the ACF-funded Head Start CARES study. Sampling frames for subsequent stages of sampling will come from the selected programs and centers. The ultimate target population to be represented by this sample at baseline is all children who are age 3 or older in the fall of 2009 and who are in their first year of Head Start.

To minimize the effects of unequal weighting on the variance of estimates, we propose sampling with probability proportional to size (PPS) in the first three stages, followed by sampling with equal probability at the final stage, with the goal of giving each child in Head Start an approximately equal chance of selection into the sample.[3] We will select 60 programs and, on average, 2 centers per program, 3 classes per center, and 12 children per class, for a total of about 4,100 children selected and about 3,400 children participating in the fall of 2009 (83 percent are expected to be eligible and to gain parental consent). At each stage of sampling, we plan to use a sequential sampling technique based on a procedure developed by Chromy.[4]

[3] At the final stage of sampling, we will be selecting a fixed number of children with equal probability within each class. This will give each child approximately the same cumulative probability of selection.

We will initially select 120 programs, and then pair adjacent selected programs within strata. (These paired programs would be similar to one another with respect to the implicit stratification variables.) We will then randomly select one from each pair to be released as part of the main sample of programs. After the initial 60 programs are selected, we will ask the Office of Head Start to call the regional ACF offices to confirm that the 60 selected programs are in good standing. If confirmed, each program will be contacted and recruited to participate in the study. If the program is not in good standing, or refuses to participate, we will release into the sample the other member of the program's pair and go through the same process of confirmation and recruitment with that program.[5] All released programs will be accounted for as part of the sample for purposes of response rates and weighting adjustments. At subsequent stages of sampling, we will release all sampled cases, expecting full participation among the selected centers and classes. At the child level, we estimate that out of 12 selected children per class, we will end up with 10 eligible children with parental consent, which is our target. We expect to lose, on average, two children per class because they are no longer enrolled, because we did not obtain parental consent, or because of the subsampling of selected siblings.

B.2. Procedures for the Collection of Information

Sampling and Estimation Procedures

Statistical Methodology for Stratification and Sample Selection. The sampling methodology is described under item B1 above. When sampling programs, we will form explicit strata using census region, metro/nonmetro status, and percentage of racial/ethnic minority enrollment. Sample allocation will be proportional to the estimated fraction of eligible children represented by the programs in each stratum. We will implicitly stratify (sort) the sample frame by other characteristics, such as percent of dual language learner (DLL) children (categorized), whether the program is a public school district grantee, and the percentage of children with disabilities. No explicit stratification will be used for selecting centers within programs, classes within centers, or children within classes, although some implicit stratification will be used. In this study, we plan to follow the selected eligible children with parental consent through their kindergarten year, as long as they complete one full year of Head Start.

Estimation Procedure. We will create analysis weights to account for variations in the probabilities of selection and variations in the eligibility and cooperation rates among those selected. For each stage of sampling (program, center, class, and child) and within each explicit sampling stratum, we will calculate the probability of selection. The inverse of the probability of selection within stratum at each stage is the sampling or base weight. The sampling weight takes into account the PPS sampling approach, the presence of any certainty selections, and the actual number of cases released. We treat the eligibility status of each sampled unit as known at each stage. Then, at each stage, we will multiply the sampling weight by the inverse of the weighted

[4] The procedure offers all the advantages of the systematic sampling approach but eliminates the risk of bias associated with that approach. The procedure makes independent selections within each of the sampling intervals while controlling the selection opportunities for units crossing interval boundaries. Chromy, J.R. "Sequential Sample Selection Methods." Proceedings of the Survey Research Methods Section of the American Statistical Association, 1979, pp. 401-406.

[5] One of the 60 sampled programs in FACES 2006 turned out to be ineligible. Three others refused to participate.

response rate within weighting cells (defined by sampling stratum) to obtain the analysis weight, so that the respondents' analysis weights account for both the respondents and nonrespondents.

Thus, the program-level weight adjusts for the probability of selection of the program and response at the program level; the center-level weight adjusts for the probability of center selection and center-level response; the class-level weight adjusts for the probability of selection of the class and class-level response; and the teacher-level weight adjusts for the probability of selection of the teacher via her classes and teacher-level response. The child-level weights adjust for the probability of selection of the child, whether parental consent was obtained, and whether various child-level data collections were obtained. The formulas below represent the various weighting steps for the cumulative weights through prior stages of selection, where P represents the probability of selection and RR the response rate at that stage of selection. Where population counts are available (such as the PIR data at the program level), we will poststratify the weights to match those counts.

$$W_{pgm} = \frac{1}{P_{pgm}} \cdot \frac{1}{RR_{pgm}}$$

$$W_{center} = W_{pgm} \cdot \frac{1}{P_{center}} \cdot \frac{1}{RR_{center}}$$

$$W_{class} = W_{center} \cdot \frac{1}{P_{class}} \cdot \frac{1}{RR_{class}}$$

$$W_{child} = W_{class} \cdot \frac{1}{P_{child}} \cdot \frac{1}{RR_{child}}$$

Degree Of Accuracy Needed for the Purpose Described in the Justification. The complex sampling plan, which includes several stages, stratification, clustering, and unequal probabilities of selection, requires the use of specialized procedures to calculate the variance of estimates. Standard statistical software assumes independent and identically distributed samples, which would indeed be the case with a simple random sample. A complex sample, however, generally has higher variances than would be calculated with standard software. We will use SUDAAN software and its Taylor Series estimation approach to produce estimates of variances that account for the complexity of the sample.

Most of the analysis will be at the child level, although some will be at the class, teacher, center, and program levels. Given various assumptions about the sample design and its impact of estimates, the sample size should be sufficiently large to detect meaningful differences. Table B.1 shows the minimum detectable differences with 80 percent power (and α=0.05) and various sample and subgroup sizes assuming an interclass correlation coefficient of 0.05 and a design effect due to unequal weighting of 1.045.

TABLE B.1

FACES 2009 MINIMUM DETECTABLE DIFFERENCES (INCLUDES HEAD START LEAVERS WHO ARE FOLLOWED)

POINT IN TIME										
	SUBGROUPS				MDD					
Time Point	Percentage in Group 1	Percentage in Group 2	Classes in Group 1	Classes in Group 2	Proportion of 0.1 or 0.9	Proportion of 0.2 or 0.8	Proportion of 0.3 or 0.7	Proportion of 0.4 or 0.6	Proportion of 0.5	Normalized Variable (mean=100, s.d.=15)
	50	50	203.5	203.5	0.090	0.120	0.138	0.147	0.150	4.054
	33	33	134.3	134.3	0.108	0.145	0.166	0.177	0.181	5.399
	33	67	134.3	272.7	0.095	0.127	0.145	0.156	0.159	4.747
Baseline	25	75	101.8	305.3	0.102	0.137	0.156	0.167	0.171	5.101
Time Point	Percentage in Group 1	Percentage in Group 2	Children in Group 1	Children in Group 2	Proportion of 0.1 or 0.9	Proportion of 0.2 or 0.8	Proportion of 0.3 or 0.7	Proportion of 0.4 or 0.6	Proportion of 0.5	Normalized Variable (mean=100, s.d.=15)
	50	50	1,649.0	1,649.0	0.046	0.061	0.070	0.074	0.076	2.280
	33	33	1,088.3	1,088.3	0.050	0.067	0.077	0.082	0.084	2.508
	33	67	1,088.3	2,209.7	0.047	0.062	0.072	0.076	0.078	2.340
Base-line	25	75	824.5	2,473.5	0.049	0.065	0.074	0.079	0.081	2.430
Kinder-garten	50	50	1100.5	1100.5	0.050	0.067	0.076	0.082	0.083	2.501
PRE-POST DIFFERENCE ESTIMATES										
	TIME POINTS				MDD					
Time 1	Time 2	Subgroup at Both Times	Children at Time 1	Children at Time 2	Proportion of 0.1 or 0.9	Proportion of 0.2 or 0.8	Proportion of 0.3 or 0.7	Proportion of 0.4 or 0.6	Proportion of 0.5	Normalized Variable (mean=100, s.d.=15)
	Spring 2010	100	3,298.0	2,959.0	0.030	0.041	0.046	0.050	0.051	1.521
		100	3,298.0	2,201.0	0.034	0.046	0.052	0.056	0.057	1.712
		50	1,649.0	1100.5	0.039	0.052	0.060	0.064	0.065	1.953
Baseline	Kindergarten	33	1,088.3	726.3	0.044	0.058	0.066	0.071	0.073	2.174

Note: Conservative assumption of no covariance for point-in-time subgroup comparisons. Covariance adjustment made for pre-post difference (Kish, p.462, Table 12.4.II, Difference with Partial Overlap). Assumes α =.05, .80 power (60 programs), ICC (rate of homogeneity) =.05, and design effect due to unequal weighting of 1.045.

MDD = minimum detectable difference; s.d. = standard deviation.

For point-in-time estimates, we are making the conservative assumption that there is no covariance between estimates for two subgroups, even though the observations may be in the same classes, centers, and/or programs. By conservative, we mean that smaller differences than those shown will likely be detectable. We show detectable differences assuming a clustering effect based on an intraclass correlation coefficient (ICC) of 0.05. We assume a similar ICC effect at the class level. This assumption is also conservative because the ICC is likely to be smaller at the class level than at the child level. For pre-post estimates, we do assume covariance between the estimates at two points in time. Evidence from another survey shows expected correlations between fall and spring estimates of about 0.5. Using this information, we applied another design effect component to the variance of estimates of pre-post differences to reflect the fact that it is efficient to have many of the same children or classes at both time points.

The top section of Table B.1 (labeled "POINT IN TIME") shows the minimum differences that would be detectable for point-in-time (cross-sectional) estimates at the class and child levels. In addition to the design effect attributable to clustering, the table incorporates the design effect attributable to unequal weighting. The bottom section (labeled "PRE-POST DIFFERENCES ESTIMATES") shows detectable pre-post difference estimates at the child level. The leftmost columns ("SUBGROUPS" and "TIME POINTS") show several sample subgroup proportions (for example, a comparison of male children to female children would be represented by "50, 50"). The class estimates all assume a sample size of 407. The child-level estimates represent three scenarios: (1) all children in fall 2009 who complete the child assessment (n = 3,298), (2) all children in spring 2010 with a completed child assessment (n = 2,959), and (3) all children in kindergarten year who complete the child assessment (n = 2,201). For example, the n = 3,298 row within the "33, 67" section represents a subgroup comparison involving both cohorts at the beginning of data collection for two subgroups, one representing one-third of that sample and the other representing the other two-thirds of that sample.

The last six columns ("MDD") show various types of variables from which an estimate might be made; the first five are estimates in the form of proportions, and the last is an estimate for a normalized variable (such as an assessment score) with a mean of 100 and standard deviation of 15. The numbers for a given row and column show the minimum differences between the two subgroups that would be detectable for a given type of variable.

At the child level, if we compared normalized assessment scores with an intracluster correlation of 0.05, a sample size of 3,298 children in fall 2009, and two approximately equal-sized subgroups (such as boys and girls, or 3-year-old and 4-year-old cohorts), our design would allow us to detect a minimum difference of 2.28 points with 80 percent power. If we compared these two subgroups again in their kindergarten year, our design would allow us to detect a minimum difference of 2.50 points.

If we performed a pre-post comparison (say, fall 2009 to kindergarten) for the same normalized assessment measure (with ICC = 0.05) for both cohorts (n = 3,298 in fall; n = 2,201 in kindergarten), we would be able to detect a minimum difference of 1.71 points. If we performed the same pre-post comparison for a subgroup representing one-half of the entire sample (n = 1,649 in fall 2009; n = 1100.5 in kindergarten), we would be able to detect a minimum difference of 1.95 points.

While not shown, the differences that would be detectable when analyzing each of the two age cohorts separately would be fairly comparable to those shown here for the mixed-age estimates. The sample size for a specific age cohort would be smaller than for both age cohorts combined, reducing the power but also reducing the design effect attributable to clustering.

Unusual Problems Requiring Specialized Sampling Procedures. We do not anticipate any unusual problems that require specialized sampling procedures.

Any use of Periodic (Less Frequent than Annual) Data Collection Cycles to Reduce Burden. We do not plan to collection data any less frequently than once per year to reduce burden.

Data Collection Procedures

As in the four past cohorts of FACES, we propose to collect data from several sources: Head Start children and their parents, Head Start teachers, program directors, center directors, and education coordinators, and kindergarten teachers. The data collection procedures are unchanged from those used in FACES 2006. Table A.1 shows the data collection waves, time periods, and methods.

The field data collection period is eight weeks long, beginning in September for the fall 2009 data collection and March for the spring follow-up waves. MPR's FACES coordinators, in conjunction with the OSC, will schedule the data collection week based on the Head Start program's availability. A maximum of eight sites will be scheduled each week. Three weeks prior to the assigned data collection week, parent interviewing will begin by telephone.

Data collection instruments are included in Appendices C through K. The instruments that will be used in FACES 2009 are for the most part the same as those used in FACES 2006. Appendix L summarizes the differences in the instruments.

Parent Interviews. On average, each parent interview is approximately 45 minutes in length. Interviews will be conducted using computer-assisted technology. We anticipate completing approximately 60 percent of the parent interviews by telephone at each wave. The remainder of the parent interviews will be completed in person, during the weeklong field data collection visit.

We will send parents an advance letter approximately four weeks prior to the start of data collection (Appendix B) to remind them about the interview, which they can complete by telephone or in person during the data collection week. Telephone interviewing will begin three weeks prior to the scheduled field visit week. During the Head Start data collection waves, in-person parent interviews will be conducted onsite at the child's Head Start center during the scheduled data collection week. When the sampled children move on to kindergarten, in-person parent interviews will be conducted in the children's homes. Onsite coordinators and field interviewers will be responsible for scheduling these interviews prior to the scheduled data collection visit.

Parents of children who are found to have left Head Start after their first year, but have not yet gone to kindergarten, will receive a shorter, 15-minute interview. We expect to locate and complete 70 percent of these interviews.

Child Assessments. Child assessments will be conducted at each data collection wave during the scheduled data collection week. During the Head Start year, child assessments will be scheduled onsite at the Head Start center by the onsite coordinator. Parents will be reminded of the child assessments in their advance letter. During the kindergarten year, child assessments will be conducted in their homes, scheduled by the field interviewer prior to and during the data collection week. On average, child assessments take approximately 45 minutes. Child assessments will be administered using computer-assisted personal interviewing. We anticipate completing assessments for 90 percent of the sampled children who remain in Head Start, and 80 percent of those who have left Head Start after the first year and are attending kindergarten.

Head Start Program Director Interview. The Head Start program director interview is a one-time, semi-structured, 30-minute telephone interview conducted by the FACES coordinator prior to the field data collection. We expect to complete 95 percent of these interviews.

Center Director and Education Coordinator Interviews. These 30-minute interviews are each administered once, during the fall 2009 data collection wave. Field interviewers administer the interviews in person, on paper. The survey items have not changed from FACES 2006. We expect to complete 95 percent of these interviews.

Head Start Teacher Interviews. These 30-minute long interviews are administered during the onsite data collection week by a member of the field staff. The OSC will schedule the interviews with the teachers in advance of the field team arrival. Head Start teacher interviews are administered using computer-assisted personal interviewing. We anticipate completing 90 percent of these interviews at each wave.

Head Start Teacher Child Report (TCR). Head Start teachers will be asked to complete a TCR for each FACES child in their classroom. Letters containing an Internet web address, login id, and password for completing the TCRs online will be sent to teachers of FACES children prior to the onsite field visit (see Appendix B for letter). During the onsite field visit, field interviewers will have hard copies of the TCR forms for teachers who would prefer to complete the forms with paper and pencil. Each TCR is expected to take approximately 10 minutes to complete. In the first wave of data collection, we anticipate teachers will have approximately 12 FACES children in each classroom, with children dispersing among other classes and teachers in later waves. The TCR forms remain essentially the same as the FACES 2006 forms. Based on experience with FACES 2006, we expect 75 percent of the TCR forms will be completed by web. We expect an overall response rate of 90 percent of TCR forms complete.

Kindergarten Teacher Questionnaire and Teacher Child Report (TCR). The kindergarten teacher questionnaire is also 30 minutes long, and is followed by a teacher child report form for each FACES child in the teacher's class. The questionnaire and TCRs are self-administered either on the web or hardcopy. Once kindergarten teachers are identified for each kindergarten child, a letter will be sent to the teacher providing an Internet web address, login id, and password for completing the survey and the TCRs online. A hardcopy version of the

questionnaire and TCRs will also be included in the letter for teachers who prefer a paper and pencil method. We estimate that approximately 60 percent of completed questionnaires and TCRs will be completed online. We expect an overall response rate of 70 percent for kindergarten teacher questionnaires. Kindergarten teacher questionnaires and TCRs will not be sent to the teachers of children who leave Head Start and do not immediately go on to kindergarten.

B.3. Methods to Maximize Response Rates and Deal with Nonresponse

There is an established, successful record of gaining program cooperation and obtaining high response rates with center staff, children, and families in research studies of Head Start, Early Head Start, and other preschool programs. To achieve high response rates, the procedures that have worked well on FACES 2006 will be continued. Marginal response rates for FACES 2006 ranged from 92 percent to 99 percent across instruments. Head Start program staff and families will be motivated to participate because they are vested in the success of the program. ACF will send a letter to selected programs, signed by Maria Woolverton, the federal project officer and a member of the senior staff at the Office of Head Start, describing the importance of the study, outlining the study goals, and encouraging their participation. Section A.9 of this submission discusses compensation payments to be made to programs, payments to parents and teachers for completion of survey instruments, and gifts to children for participating in the assessments. All of these, which we have used in FACES 2006, will help ensure a high level of participation. Obtaining the high response rate we expect to attain makes the possibility of nonresponse bias less likely, which in turn makes our conclusions more generalizable to the Head Start population.

Families that choose to leave the Head Start program prior to enrolling in kindergarten will complete a short interview by telephone and their children will be administered the same assessments as other children in the study. Because these families may be difficult to track down for a telephone interview, we will use specialized locating resources, ranging from calling contacts the respondent listed in prior interviews to directory assistance, and database searchers such as LexisNexis and Accurint.

We will calculate both unweighted and weighted, marginal and cumulative, response rates at each stage of sampling and data collection. As reflected in the American Association for Public Opinion Research (AAPOR) industry standard for calculating response rates, the numerator of each response rate will include the number of eligible completed cases and the denominator will include the number of eligible selected cases.

B.4. Test of Procedures or Methods to be Undertaken

Most of the scales and items in the proposed parent interview, child assessment, program director interview, and teacher interviews have been successfully administered in FACES 2006.

B.5. Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing Data

The team is lead by Maria Woolverton, project officer; Dr. Jerry West, project director; Dr. Louisa Tarullo, principal investigator; Ms. Cassandra Meagher, survey director; and Ms. Annalee Kelly, deputy survey director. The plans for statistical analyses for this study were developed by Mathematica Policy Research, Inc. (MPR) with support provided by Juarez and Associates and The Educational Testing Service. Additional staff consulted on statistical issues at MPR include Mr. John Hall and Ms. Barbara Carlson, senior statisticians, and Dr. Don Rock, a consultant to MPR on psychometric issues.