Head Start Family and Child Experiences Survey (FACES 2014–2018) OMB Supporting Statement for Recruitment of Programs and Selecting Centers

Part B: Collection of Information Involving Statistical Methods

January 16, 2014

CONTENTS

TABLES

FIGURES

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

The Office of Planning, Research and Evaluation (OPRE), Administration for Children and Families (ACF), U.S. Department of Health and Human Services (HHS), is proposing to collect data for a new round of the Head Start Family and Child Experiences Survey (FACES). FACES 2014–2018 features a new Core Plus study design that consists of two Core studies—the Classroom + Child Outcomes Core and the Classroom Core—and Plus studies to include additional survey content of policy or programmatic interest to be determined. The Classroom + Child Outcomes Core, occurring during the 2014-2015 program year, collects child-level data along with program and classroom data from a subset of programs while other programs will only have data collected on program and classroom information (see Part A for details). In spring 2017, the Classroom Core will be conducted focusing on program and classroom data collection only for all programs.

ACF, contracting with Mathematica Policy Research (Mathematica) and its subcontractors, Juárez and Associates and Educational Testing Service, requests permission to contact 230 Head Start programs that will be selected to participate in FACES 2014–2018 (180 in the Core studies and up to 50 additional programs for the Plus studies) for the purpose of gathering information for sampling centers, classrooms, and children. In this package, we present the sampling plans for all these levels and the procedures for recruiting programs and selecting centers in 2014 and contacting them again in 2016. A separate package will be submitted to request clearance for the FACES 2014–2018 data collection, including selecting classrooms and children for the study and gathering consent for children, data collection instruments and procedures, data analyses, and the reporting of study findings.

B.1. Respondent Universe and Sampling Methods

The target population for FACES 2014–2018 is all Head Start programs in the United States, their classrooms, and the children and families they serve. The sample design is similar to the one used for FACES 2009 in some respects, but with some key differences noted below. FACES 2014–2018 will use a stratified multistage sample design with four stages of sample selection: (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) for a random subsample of programs, children within classes. To minimize the burden on parents/guardians who have more than one child selected for the sample, we will also randomly subsample one selected child per parent/guardian, a step that was introduced in FACES 2009.

The frame that will be used to sample programs is the 2012–2013 Head Start Program Information Report (PIR), which is an updated version of the frame used for previous rounds of FACES. We will exclude from the sampling frame: Early Head Start programs, programs in Puerto Rico and other U.S. territories, migrant and seasonal worker programs, programs that do not directly provide services to children in the target age group, programs under transitional management, and programs that are (or will soon be) defunded.[1] The center frame will be

[1] We will work with the Office of Head Start (OHS) to update the list of programs before finalizing the sampling frame. Grantees and programs that were known by OHS to have lost their funding or otherwise closed between summer 2013 and winter 2014 will be removed from the frame, and programs associated with new grants awarded since then will be added to the frame.

developed through contacts with the sampled programs. Similarly, the classroom and child frames will be constructed after centers and classroom samples are drawn. All centers, classrooms, and children in study-eligible, sampled programs will be included in the center, classroom, and child frames, respectively, with two exceptions. Classrooms that receive no Head Start funding (such as prekindergarten classrooms in a public school setting that also has Head Start-funded classrooms) are ineligible. Also, sampled children who leave Head Start between fall and spring of the program year become ineligible for the study. Sampling of centers, classrooms, and children, which we describe below, are not a part of information-gathering activities for which clearance is being requested in this submission.

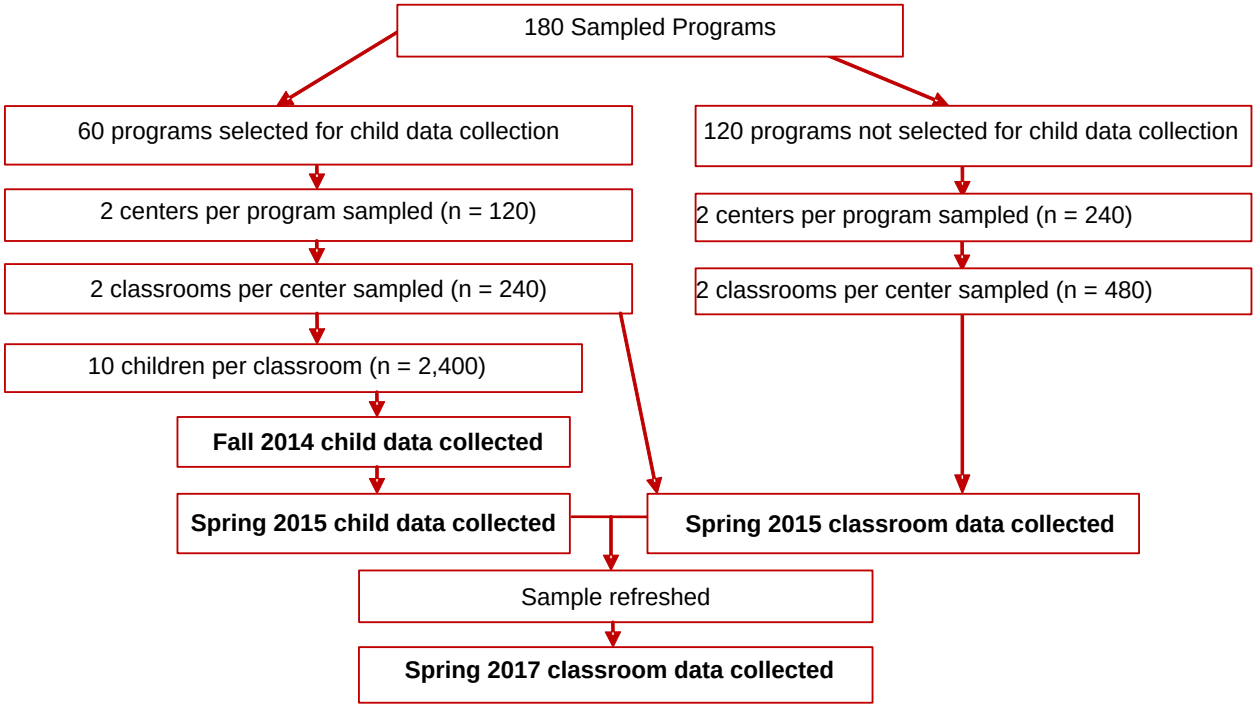
The sample design for the new round of FACES is based on the one used for FACES 2009, which was based on the designs of the four previous rounds. But unlike the earlier rounds of FACES, the sample design for FACES 2014–2018 will involve sampling for two newly designed study components: Classroom + Child Outcomes Core and Classroom Core. The Classroom + Child Outcomes Core study will involve sampling at all four stages (programs, centers, classrooms, and children) and the Classroom Core study will involve sampling at the first three stages only (excluding sampling of children within classes). Under this design, the collective sample size across the two studies will be larger than in prior rounds of FACES at the program, center, and classroom levels, allowing for more powerful analyses of program quality, especially at the classroom level. Also new to the FACES 2014–2018 design, the child-level sample will represent children enrolled in Head Start for the first time and those who are attending a second year of Head Start. This will allow for a direct comparison of first- and second-year program participants and analysis of child gains during the second year. Previously, FACES followed newly enrolled children through one or two years of Head Start and then through spring of kindergarten. FACES 2014–2018 will follow the children only through the fall and spring of one program year.

To minimize the effects of unequal weighting on the variance of estimates, we propose sampling with probability proportional to size (PPS) in the first two stages. At the third stage we will select an equal probability of classrooms within each sampled center and, in centers where children are to be sampled, an equal probability sample of children within each sampled classroom. The measure of size for PPS sampling in each of the first two stages will be the number of classrooms. This sampling approach maximizes the precision of classroom-level estimates and allows for easier in-field sampling of classrooms and children within classrooms. We will select a total of 180 programs across both Core study components. Sixty of the 180 programs sampled for the Core study will be randomly subsampled with equal probability within strata to be included in the Classroom + Child Outcomes study. Within these 60 programs, we will select, if possible, two centers per program, two classes per center, and a sufficient number of children to yield 10 consented children per class, for a total of about 2,400 children at baseline.

Based on our experience with earlier rounds of FACES, we estimate that 70 percent of the 2,400 baseline children (about 1,680) will be new to Head Start. We expect a program and study retention rate of 90 percent from fall to spring, for a sample of 2,160 study children in both fall 2014 and spring 2015, of which about 1,512 (70 percent) are estimated to have completed their first Head Start year.

The Classroom Core study component will include the 60 programs where students are sampled plus the remaining 120 programs from the sample of 180. We will select, from the additional 120 programs, two centers per program, and two classrooms per center. Across both study components, we will have a total of 360 centers and 720 classrooms for spring 2015 data collection. For follow-up data collection in spring 2017, we will select a refresher sample[2] of programs and their centers so that the new sample will be representative of all programs and centers at the time of follow-up data collection, and we will select a new sample of classrooms in all centers. Figure B.1 is a diagram of the sample selection and data collection procedures. At each sampling stage, we will use a sequential sampling technique based on a procedure developed by Chromy.[3]

Figure B.1. Flow of Sample Selection Procedures



For the Core studies, we will initially select 360 programs, and 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 180 programs are selected, we will ask the Office of Head Start (OHS) to confirm that the 180 selected programs are in good standing. If confirmed, each program will be contacted and recruited to participate in the study: the 60 programs subsampled for the Classroom + Child Outcomes Core will be recruited in spring 2014 (for fall 2014 participation); the remaining 120 programs will be recruited in fall

[2] The process of "freshening" a sample of students has been used for many NCES longitudinal studies. The freshening of the program sample for FACES 2014-2018 will use well-established methods that ensure that the refreshed sample can be treated as a valid probability sample.

[3] 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,* Alexandria, VA: American Statistical Association, 1979, pp. 401–406.

2014 (for spring 2015 participation). 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. All released programs will be accounted for as part of the sample for purposes of calculating 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 parental consent was not granted, or because of the subsampling of selected siblings.

We will select centers PPS within each sampled program using the number of classrooms as the measure of size, again using the Chromy procedure. For the Classroom + Child Outcomes Core, we will randomly select classrooms within center with equal probability. Classrooms with very few children will be grouped with other classrooms in the same center for sampling purposes to ensure a sufficient sample yield.[4] Once classrooms are selected, we will select an equal probability sample of 12 children per class, with the expectation that 10 will be eligible and will receive parental consent.

B.2. Procedures for Collecting 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 classrooms represented by the programs in each stratum.[5] We will implicitly stratify (sort) the sample frame by other characteristics, such as percentage 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 (such as the percentage of children who are dual language learners) may be used for center selection.

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] If the number of children per class is not available at the time of classroom sampling, we will randomly sample three classrooms, and then randomly subsample two for initial release. If these two classrooms are not likely to yield 20 children, we will release the third classroom as well.

[5] We will stochastically round the stratum sizes as needed.

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; and the class-level weight adjusts for the probability of selection of the class and class-level response. The child-level weights adjust for the subsampling probability of programs for the Classroom + Child Outcomes Core; the probability of selection of the child within classroom, whether parental consent was obtained, and whether various child-level instruments (for example, direct child assessments and parent surveys) 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. Because FACES 2014-2018 includes all children (not just those newly enrolled), we will post-stratify to know totals at each weighting stage.

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

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

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

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

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 using 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 larger variances than would be calculated with standard software. Two approaches for estimating variances under complex sampling, Taylor Series and replication methods, can be estimated by using SUDAAN and special procedures in SAS, Stata, and other packages.

Most of the analyses will be at the child and classroom levels. Given various assumptions about the sample design and its impact of estimates, the sample size should be sufficiently large to detect meaningful differences. In Table B.1, we show the minimum detectable differences with 80 percent power (and =0.05) and various sample and subgroup sizes, assuming different intraclass correlation coefficients for classroom- and child-level estimates at the various stages of clustering (see table footnote).

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. 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 Subgroup Comparisons") shows the minimum differences that would be detectable for point-in-time (cross-sectional) estimates at the class and child levels. We have incorporated the design effect attributable to clustering. The bottom section (labeled "Estimates of Program Year Gains") shows detectable pre-post difference estimates at the child level. Examples are given below.

The columns farthest to the left ("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 child-level estimates represent two scenarios: (1) all consented children in fall 2014 (n = 2,400) and (2) all children in spring 2015 who remained in Head Start (n = 2,160). For example, the n = 2,400 row within the "33, 67" section represents a subgroup comparison involving children at the beginning of data collection for two subgroups, one representing one-third of that sample (for example, children in bilingual homes), the other representing the remaining two-thirds (for example, children from English-only homes).

The last few columns ("MDD") show various types of variables from which an estimate might be made; the first two are estimates in the form of proportions, the next is an estimate for a normalized variable (such as an assessment score) with a mean of 100 and standard deviation of 15 (for child-level estimates only), and the last shows the minimum detectable effect size—the MDD in standard deviation-sized units. The numbers for a given row and column show the minimum underlying differences between the two subgroups that would be detectable for a given type of variable with the given sample size and design assumptions.

Table B.1. FACES 2014–2018 Minimum Detectable Differences

POINT IN TIME SUBGROUP COMPARISONS								
	SUBGROUPS				Minimum Detectable Difference			
Time Point	Percentage in Group 1	Percentage in Group 2	Classes in Group 1	Classes in Group 2	Proportion of 0.1or 0.9	Proportion of 0.5		Minimum Detectable Effect Size
Spring 2015	50	50	360	360	.084	.140		.280
	33	67	238	482	.090	.149		.298
	15	85	108	612	.119	.198		.392
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 0.5	Normalized Variable (Mean = 100, s.d.= 15)	Minimum Detectable Effect Size
Fall 2014	50	50	1,200	1,200	.072	.119	3.578	.239
	33	67	792	1,608	.076	.127	3.805	.254
	40	30	960	720	.087	.144	4.321	.288
Spring 2015	50	50	1,080	1,080	.072	.121	3.617	.241
ESTIMATES OF PROGRAM YEAR GAINS								
TIME POINTS					Minimum Detectable Difference			
Time 1	Time 2	Percent Subgroup at Both Times	Children at Time 1	Children at Time 2	Proportion of 0.1 or 0.9	Proportion of 0.5	Normalized Variable (Mean = 100, s.d.= 15)	Minimum Detectable Effect Size
Fall 2014	Spring 2015	100	2,400	2,160	.038	.063	1.887	.126
		70	1,680	1,512	.045	.075	2.255	.150
		40	960	864	.060	.100	2.983	.199

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 =.05 (two-sided), .80 power. For classroom-level estimates, assumes 180 programs, 360 centers, between-program ICC=.2, between-center ICC = .2. For child-level estimates, assumes 60 programs, 120 centers, between-program ICC = .05, between-center ICC = .05, between-classroom ICC = .05.

Note: The minimum detectable effect size is the minimum detectable difference in standard deviation-sized units.

If we were to compare two equal-sized subgroups of the 720 classrooms in spring 2015, our design would allow us to detect a minimum difference of .280 standard deviations with 80 percent power. At the child level, if we were to compare normalized assessment scores with a sample size of 2,400 children in fall 2014, and two approximately equal-sized subgroups (such as boys and girls), our design would allow us to detect a minimum difference of 3.578 points with 80 percent power. If we were to compare these two subgroups again in the spring of 2015, our design would allow us to detect a minimum difference of 3.617 points.

If we were to perform a pre-post comparison (fall 2014 to spring 2015) for the same normalized assessment measure, we would be able to detect a minimum difference of 1.887 points. If we were to perform the same pre-post comparison for a subgroup representing 40 percent of the entire sample (n = 960 in fall 2014; n = 864 in spring 2015), we would be able to detect a minimum difference of 2.98 points.

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 reduce burden by collecting data less frequently than once per year.

Data Collection Procedures

Head Start programs will be selected in late winter 2014. Upon OMB approval of this information collection request, a letter signed by the Acting Director of the Office of Head Start will be sent to each sampled program's director. The letter will describe the study goals and the importance of the study and introduce the Mathematica team who will be doing the study on ACF's behalf (Part A, Appendix A-1). A letter and study fact sheet will be sent by Mathematica following the introductory letter (Part A, Appendix A-2 and A-3).

Program directors will then receive a phone call from a member of the study team to answer any questions about the study. Using a prepared script (Attachment 1), the study team will review our request for information and ask for information about centers (names, addresses, and estimated enrollment), how services are organized (center-based, home-based, combination, or locally designed), and scheduling specifics (hours of operation, program year start and end dates, and full- versus part-day program). This information will be recorded for use in preparing the data collection plans for the study programs. Directors will also be asked to identify a staff member to serve as an on-site coordinator who will work with the study team to recruit participants, develop a data collection plan, and help schedule site visits.

As another element of this information collection, a member of the study team will call on-site coordinators (after they have receiving a letter describing the study; see Part A, Appendix A-4) and, using a prepared script (Attachment 2), ask the coordinator to provide or confirm information about the centers in their program. Finally, once centers are selected, center directors will receive a letter (Part A, Appendix A-5) along with the study fact sheet.

B.3. Methods to Maximize Response Rates and Data Reliability

We do not anticipate problems contacting and gaining the cooperation of Head Start programs, and in gathering information from program directors and on-site coordinators. The study team will conduct calls with program directors and on-site coordinators during business hours, at times that coincide best with their schedules. We will use the same approach and procedures that were used successfully in FACES 2006 and FACES 2009. Program response rates in both rounds exceeded 95 percent.

B.4. Test of Procedures or Methods

The proposed procedures were used successfully in FACES 2009, and there are no plans to test the procedures.

B.5. Individuals Consulted on Statistical Methods

The team is led by Maria Woolverton, Federal Contracting Officer's Representative (COR); Dr. Jerry West, project director; Drs. Louisa Tarullo and Nikki Aikens, co-principal investigators; and Annalee Kelly, survey director. Additional staff consulted on statistical issues include Barbara Carlson, a senior statistician at Mathematica, and Dr. Margaret Burchinal, a consultant to Mathematica on statistical and analytic issues.