Head Start Family and Child Experiences Survey

FACES 2014

Self-Guided Training

Prepared by

Jerry West, Ph.D.

University of Maryland, College Park

For Child Care and Early Education Research Connections,
National Center for Children in Poverty, Columbia University

With support from the Office of Planning, Research and Evaluation,
Administration for Children and Families,
U.S. Department of Health and Human Services

2018

The development of these training materials benefited greatly from comments and suggestions provided by the staff of the Office of Planning, Research and Evaluation. Thanks is also extended to students at the graduate program of the College of Education, University of Maryland at College Park for their feedback and to participants in the summer data training program at the Inter-university Consortium for Political and Social Research, University of Michigan, who pilot tested the training. Any errors remain the responsibility of the Research Connections project.
# TABLE OF CONTENTS

Introduction ................................................................................................................................................... 1

Module 1:  Introducing FACES 2014 ........................................................................................................ 4

Module 2.1: FACES 2014 Core Study Instruments ................................................................. 50

Module 2.2: Faces 2014 Family Engagement Plus Study Instruments .......... 86

Module 3:   Data Access and Resources ...................................................................................... 104

Module 4:   Sampling Weights and Variance Estimation ....................................................... 155

Module 5:   Tips for Working with the Data .............................................................................. 210
Introduction

This online training is designed to introduce you to the Head Start Family and Child Experiences Survey, 2014 (FACES 2014). There are six training modules. They provide information on the purposes and history of FACES, introduce you to a new study design that was used for the first time in FACES 2014, describe the FACES 2014 dataset, and offer guidance on how to use and analyze the data to answer a wide range of research questions. The training is designed for a broad audience that includes researchers with little or no experience working with large-scale data and the methods required to analyze such data and those with years of experience working with these types of data and who are well versed in using the methods necessary to produce valid and reliable results.

Each module includes a set of PowerPoint slides with note pages. The slides and note pages are meant to be reviewed together. The slides include a set of bullets that outline the main or key points for a given topic (for example, what are the purposes of FACES?) and the accompanying note pages provide more description and discussion of the topic (for example, the effects of using weighted versus unweighted data). Generally, it is best to read review the bullets on the slides first and then the notes. When reading the notes, you may want to refer back to the bullets. In some cases, there may not be any notes for a specific slide. This occurs when the information on the slide is sufficient to cover the main or key points.

Each module also includes a review quiz that is designed to reinforce your understanding of the information that is presented on the slides and in the notes pages. Once you have answered all of the questions, an answer sheet is provided that contains the correct answers and rationale. Some modules include exercises that are intended to increase your understanding of a topic and for a few others you will be linked to documents for additional information about a specific topic.

Many of the topics covered in the training modules are discussed in greater detail in the FACES 2014 user's manual. Links are provided in each of the modules to sections of the FACES 2014 user's manual where you will find more information on a specific topics such as the construction and interpretation of scores derived from children’s responses to child assessment items. There are also links to various user's manual appendices such as those containing the questionnaires that were used to collect data from parents, teachers, and other Head Start staff. For other topics (for example, purposes and use of sampling weights), the material found in the training modules is more extensive than what is found in the user's manual.

Each of the six training modules covers several topics. While there is a structure and an order to how the modules and topics are introduced, you may access and complete them in any order and at your own pace. Estimates are provided of the amount of time each module is expected to take to complete.

Access to Training Material

You have two options for accessing the training material. Your first option is to access each of the six training modules and review the information included in each using PowerPoint software. You will need to view the PowerPoint document using the notes page selection. The second option is to access the same information on pdf files. Each slide and its accompanying notes are provided on a single page.
each option you will be linked directly to the review quizzes, exercises, sections and appendices in the user's manual, and any supplemental material by clicking on the links embedded in the document.

Overview of Training Modules

Below is a list of the training modules, a brief description of their contents, and an estimate of the amount of time it will take to complete each. **Click on any module listed to begin exploring FACES 2014.**

**Module 1: Introducing FACES 2014**

This module provides an introduction to FACES 2014 for researchers regardless of their expertise and experience. It provides basic information about the purposes and history of FACES as a source for national data on Head Start. It includes descriptions of the FACES 2014 design, target populations and samples, study components, methods of data collection and the data collection schedule.

*Estimated time to complete the module: 45-60 minutes*

**Module 2.1: FACES 2014 Core Study Instruments**

This module provides all researchers with an overview of the FACES 2014 Core study instruments. It describes the FACES 2014 child assessment battery and the pathways that different groups of children follow through the assessment, and the study’s survey instruments and interviews. It identifies the purposes of each instrument and its contents, and connects the instrument to the types of questions it was designed to answer. The individual items found in each of the instruments can be found in the appendices to the FACES 2014 user's manual with a few exceptions. FACES 2014 used several copyrighted instruments or scales (these include the direct child assessments, classroom observations, and indirect assessment of social skills and learning behaviors) and agreements with instrument developers and publishers sometimes did not allow individual items to be released.

*Estimated time to complete the module: 30 – 45 minutes*

**Module 2.2: FACES 2014 Family Engagement Plus Study Instruments**

This module describes the two interviews that were conducted with a sample of Head Start parents and family service staff (FSS) as part of the Family Engagement Plus study. It identifies the purposes of each interview and its contents. The supplemental questions pertaining to families’ engagement with the Head Start program that were included on the spring Core parent and teacher surveys are also described. Again, individual survey and interview items can be found in the appendices to the user’s manual. The material in this module is directed to all researchers who want to explore these data.

*Estimated time to complete the module: 20-30 minutes*

**Module 3: Data Access and Resources**

This module introduces the FACES 2014 dataset to researchers with all levels of experience and expertise. It contains the information that researchers need to work with the FACES 2014 data files. It identifies and describes the resources available to support researchers in using the data files.

*Estimated time to complete the module: 45-60 minutes*
Module 4: Sampling Weights and Variance Estimation

This module is designed to assist researchers in developing accurate and reliable answers to questions about Head Start and the children and families who are served by the program. The module includes two topics that are central to this overall goal. The first topic focuses on sampling weights. It includes a general introduction to sampling weights (what they are, how they are created and used), followed by a description of the sampling weights included with the FACES 2014 data. The second topical module focuses on estimating variances and standard errors for complex sample designs. It begins with an introduction to the issues surrounding the estimation of variances and standard errors when the data come from a complex sample design, not a simple random sample (SRS), and identifies different methods for calculating variances and standard errors for complex designs. Experienced researchers may want to opt out of those sections which introduce sampling weights and variance estimation in general. However, all researchers would benefit from reviewing those parts that deal directly with FACES 2014 sampling weights and variance estimation.

Estimated time to complete the module: 45-60 minutes

Module 5: Tips for Working with the Data

In this module, researchers are introduced to some of the technical issues and methods that they will need to be familiar with when using FACES 2014 data. Several topics are covered including merging data files, teacher-versus class-level analysis, cross-cohort analysis, and guidelines for data users. Researchers should do a cursory review of the material covered under each topic to see whether it is appropriate for their level of expertise and experience.

Estimated time to complete the module: 20-30 minutes
Head Start Family and Child Experiences Survey

FACES 2014

SELF-GUIDED TRAINING

Module 1

Introducing FACES 2014

Prepared by
Jerry West, Ph.D.

University of Maryland, College Park

For Child Care and Early Education Research Connections,
National Center for Children in Poverty, Columbia University

With support from the Office of Planning, Research and Evaluation,
Administration for Children and Families,
U.S. Department of Health and Human Services

The development of these training materials benefited greatly from comments and suggestions provided by the staff of the Office of Planning, Research and Evaluation. Thanks is also extended to students at the graduate program of the College of Education, University of Maryland at College Park for their feedback and to participants in the summer data training program at the Inter-university Consortium for Political and Social Research, University of Michigan, who pilot tested the training. Any errors remain the responsibility of the Research Connections project.
Module 1 introduces researchers, regardless of their level of expertise and experience, to FACES 2014. It begins by describing the purpose of FACES and its history as a source of national level data on the population of children and families served by the Head Start program. It identifies the types of questions that can and cannot be answered using the FACES data and notes other limitations of the design. Next, it introduces FACES’ new design – Core Plus Study design- that was used for the first time with FACES 2014, and which addresses some of the limitations of prior rounds of FACES.

Key features of this design are described and compared to earlier FACES designs. Module 1 describes the FACES 2014 sample, the instruments (for example, child assessment, parent survey and teacher child reports) that are used to collect data from different sources (for example, children, parents, and teachers), and the data collection schedule that includes fall 2014 and spring 2015.
Topic 1.1
Purpose and History of FACES
Successive rounds of FACES, in 1997, 2000, 2003, 2006, 2009, and now FACES 2014 have served as a rich source of national level data on Head Start programs and the children and families served by Head Start. FACES data have been used to report specifically on key indicators, including child performance in vocabulary, letter knowledge, and math, as well as teacher reports on children’s social skills, parent reports on child health and parent-child reading, and observations of teacher-child interaction in Head Start classrooms.

Data from recurring rounds of FACES have been used to examine changes in Head Start programs and services and in the population of children who attend Head Start and their families.
FACES Is Not a Study of Head Start Impacts

- Cannot answer questions about whether Head Start works or whether participation in Head Start has an impact on child and family outcomes
- No control or comparison group
- Study sample includes only
  - Head Start programs and centers
  - Head Start classrooms
  - Children who are enrolled in Head Start and their families

FACES is a descriptive study of Head Start and is not designed to answer questions about whether Head Start works or whether participation in Head Start has an impact on child and family outcomes. There is no control or comparison group. FACES sample is limited to Head Start programs, centers, and classrooms, and to the children and their families participating in the Head Start program.
FACES data can be used to answer questions about Head Start programs and centers, such as what are their structural characteristics, policies and practices? FACES data can also be used to answer questions about structural and process quality of Head Start classrooms, and about the characteristics and experiences of Head Start teachers and management staff.

FACES is designed to provide rich descriptive information about the children who attend Head Start and their families at key points in time such as at program entry, at the end of the Head Start year and after children have left Head Start and are attending kindergarten (FACES 1997-2009 only).

FACES data are used to answer questions about whether and how children’s skills, family life, and program services and educational experiences change over the Head Start year and after they have graduated from Head Start (FACES 1997-2009 only). For example, researchers can use data from a single round of FACES (such as FACES 2009) to answer questions about how much did children’s language and math skills improve from fall to spring of the 2009-2010 program year and after one or two years in Head Start.

Because many of children’s skills and other program, child and family characteristics are measured the same at each round of FACES, it is possible to compare the children and families who are served by Head Start over the past 15+ years (children enrolled in Head Start in 1997-2014).
Although FACES is not intended as an evaluation of the impact of Head Start, it can be used to study differences in children’s and families’ Head Start experiences and how these differences are related to a range of child and parent/family outcomes. This slide lists examples of the types of analytic questions that can be answered with FACES data.

The longitudinal design enables researchers to examine children’s development over the Head Start year. Differences in children’s development can be related to a host of child and family characteristics, as well as characteristics of programs and program services.

The inclusion of classroom observation measures and teacher reports of classroom activities allows researchers to describe the quality of Head Start classrooms and instruction and the relationship of classroom quality and classroom activities to child outcomes.

Data from classroom observations and teacher surveys can be used to explore the associations between different domains and dimensions of quality and characteristics of classroom teachers, their teaching experience, and the student composition of the Head Start classrooms.
From the time it was launched in 1997 through the 2009 cohort, the design of FACES stayed much the same. A new nationally representative sample of programs in the 10 geographically-based Head Start regions nationwide (Head Start Regions I-X) and the children attending those programs was selected every three years. Tribal Head Start programs (Head Start Region XI) and programs serving the children of migrant and seasonal workers (Head Start Region XII) were not eligible for FACES. Each FACES sample included children who were 3- or 4-years old and entering Head Start for the first time. Children who were attending Head Start for a second year, primarily 4-year olds, were not eligible for FACES.

Children in the sample were followed longitudinally with three (primarily 4-year olds) or four waves (primarily 3-year olds) of data collection depending on whether they attended one or two years of Head Start before moving on to kindergarten (FACES 1997 followed children to first grade).
This table shows the general data collection schedule that was used in FACES from 1997 through 2009; although there have been some deviations from this schedule (for example, FACES 2000 and 2003 observed classrooms in both the fall and spring in the first year of the study when all children were attending Head Start). During the first year of the study, data are collected in both the fall and spring and in the spring only thereafter.

At each wave of data collection, children were assessed, their parents were interviewed, and their teachers reported on their social and emotional development, academic accomplishments and special needs (Teacher Child). At each wave, teachers were asked questions about themselves and the classes they taught. Program and center directors and education coordinators were surveyed only once and during the first year of each new cohort.
The design that was used for FACES 1997-2009 had a number of strengths, many of which are listed here. FACES large and diverse sample is used to produce accurate estimates of key characteristics of the Head Start population and subgroups within that population (for example, Hispanic/Latino children and families, children who entered the program at age 3 and at age 4 and who were expected to attend the program for one or two years).

Collecting data at multiple time points produces information on children and families at particular points in time (for example, program entry and exit) and changes in children and families during the time that they are enrolled in the program and beyond. The FACES assessment battery measures children’s cognitive, language, literacy, physical and socioemotional development and surveys of their parents, teachers and program administrators provide information on the different environments that influence their development.

FACES nested design contributed to the cost effectiveness of the study and its operational feasibility, and facilitates the analysis of classroom and program impacts on child outcomes. Finally, data from a new sample of programs and children every three years can be used to study changes and trends in Head Start.
While the design of FACES 1997-2009 had many strengths, it also had limitations. Among these were small program, center and classroom samples, which made it difficult to produce accurate estimates of subgroups of programs, centers and classrooms with different characteristics and to detect subgroup differences. Also, the sample sizes for certain population groups were small, which made it difficult to develop accurate estimates of their characteristics (for example, American Indian and Alaska Native children and immigrant children).

The sample only included children who were attending Head Start for the first time and there was no fall data collection for those children in the sample who continued in Head Start for a second year. As a result, it was not possible to get a picture of the full Head Start population nor was it possible to estimate accurately the gains children made in their school readiness skills during their second year in the program. Data on critical child outcomes were available only for children who could be assessed in English or Spanish.

Data on children’s social development and on their home and school environments were provided by parent and teacher reports (i.e., there were no direct assessments of children’s social and emotional development, and there were no observation measures of children’s homes or kindergarten classrooms).
Topic 1.2
A New Study Design:
FACES 2014
There was no FACES 2012. Instead, the Administration for Children and Families (ACF) funded an extensive redesign effort. The overall goal was to ascertain what changes should be made to the design of FACES to ensure that it was responsive to the information needs of ACF and Head Start today and what could be done to improve the ability of the study to respond to new and emerging issues and question in a more timely manner. The questions listed here are those that guided the redesign effort.

**Key Questions Guiding the FACES Redesign**

- What are the most important purposes of FACES?
- What information does Head Start need to better serve children and families? For program accountability? From whom and in what form?
- How should FACES be designed in order to provide accurate and reliable information for the identified purposes?
- What changes should be made to the FACES design to increase its efficiency and its ability to respond to information needs in a more timely manner?
The design that emerged from the redesign effort shares many features with prior rounds of FACES. These are listed on the slide. Maintaining the ability to produce a national picture of Head Start, its children and families continued to be the central feature of the FACES design. Also, having individual level data about the children and families served by Head Start distinguishes FACES from the other main source of national Head Start data – Head Start Program Information Report or PIR.

The data that are collected on new Head Start samples at regular intervals can be used to study trends in the population of the children and families who are served by the program and in program policies and practices. Data from multiple sources provide a broad view of Head Start, its children and families.
The new FACES design differs in some important ways from prior rounds of FACES. Some of these are in response to limitations of the FACES 1997-2009 design (for example, FACES now includes both children enrolled for a first and a second program year and larger program and classroom samples) and others are intended to increase the efficiency of FACES (for example, all data are collected in a single program year and Head Start staff answered web surveys instead of being interviewed in person or by phone). Children will no longer be followed into kindergarten; thus, placing greater emphasis on Head Start programs and classrooms, and on children’s experiences while attending Head Start.
In Fall 2014 FACES Introduced Its New Core Plus Study Design

- Core studies
  - Describe the quality of Head Start classrooms and/or the school readiness skills of Head Start children for specific program years
  - Describe the changes or trends in the quality of classrooms and children’s outcomes over time
  - Provide key information about classrooms and programs at two-year intervals and about children and families at four-year intervals

- Plus studies allow FACES to respond flexibly to new policy and programmatic issues and questions, and address topics in greater depth

FACES new Core Plus study design was first launched in fall 2014. This design is intended to provide information on key indicators at regular intervals while at the same time enabling FACES to explore new policy and programmatic issues and topics in greater depth. The design includes Core studies that capture key characteristics and indicators at the program, classroom and child levels and Plus studies that are used to study new policy and program issues and questions and to delve deeper into topics found in the Core.
Plus studies can take different forms. The three basic types of Plus studies are listed on the slide, but there are many other possibilities for Plus studies. Topical modules may be nested in an existing FACES instrument such as the parent survey or teacher survey. Special studies may often, but not always involve sampling a new population or oversampling a population to increase its numbers. Design studies are used to conduct methodological inquiries and to pretest measures before they are introduced into the FACES design.
FACES 2014 conducted two Core studies that differ from each other in purpose and design. The Classroom Core is designed to answer questions about Head Start programs, staff and classrooms. The Classroom + Child Outcomes Core is designed to answer questions about the children and families served by Head Start.

FACES 2014 also fielded three Plus studies. The data from only one of the three is included on the FACES 2014 data files currently archived at Research Connections – Family Engagement Plus study. It was designed to provide detailed information about the engagement and service provision experiences of Head Start families as well as information about family service staff who have not traditionally been a part of FACES.

The 5E-Early Ed educator survey was piloted in spring 2015. The goal was to provide the developers of the instrument, which was adapted from a version used with elementary school teachers, with information on the performance of the items and scales when administered to a sample of preschool teachers. Depending on the outcome of the pilot, the measure may be used in future rounds of FACES.

In recognition of the need for data on the children and families served by these programs, a first-ever study of Region XI Head Start was designed and conducted during the 2015-2016 program year.

More details on each of the Core and Plus studies are found on the following slides and notes.
The Classroom Core is designed to answer questions about Head Start programs, staff and classrooms and includes a larger sample of programs and classrooms than prior rounds of FACES. Data collection is limited to classroom observations and staff surveys (program and center directors and classroom teachers). All data collection occurs in the spring of the program year. The Classroom Core is repeated every two years with the next round scheduled for spring 2017. It collects no child level data.

NOTE: The sample sizes shown here are the expected sample sizes. The actual number of programs, centers, and classrooms that participated in FACES 2014 are described in Topic 1.3 – FACES 2014 Sample.
The Classroom + Child Outcomes Core is designed to answer questions about the children and families served by Head Start. It is conducted in a subsample of the Head Start programs selected for the Classroom Core and includes child-level data collection. In addition to the spring classroom observations and staff surveys that occur under the Classroom Core, it includes direct child assessments, parent surveys and teacher reports of children’s development, collected in both the fall and spring of the program year.

The Classroom + Child Outcomes Core is repeated every four years and the next sample of Head Start children will be selected and assessed in fall 2018.

NOTE: The sample sizes shown here are the expected sample sizes. The actual number of programs, centers, classrooms and children that participated in FACES 2014 are described in Topic 1.3 – FACES 2014 Sample.
The Family Engagement study was designed to provide detailed information about the engagement and service provision experiences of Head Start families as well as information about family service staff who have not traditionally been a part of FACES. The study was conducted in the 60 programs with child-level data collection in spring 2015. It included qualitative interviews with a subsample of parents whose children participated in FACES and with a sample of family service staff. Supplemental item sets were also included in the spring Core parent survey and teacher survey.

**FACES 2014 Family Engagement Plus Study**

- Provides information about:
  - Engagement and service provision experiences of Head Start families
  - Direct providers of services to parents and families whose voices have not been captured in earlier rounds of FACES
- Conducted in 60 programs with child-level data collection
- Data collected in spring 2015
- One-hour qualitative interviews with subsample of parents with children participating in FACES and sample of family service staff
- Supplemental items added to the Core parent and teacher surveys
The 5E-Early Ed educator survey was piloted in spring 2015. The goal was to provide the developers of the instrument, which was adapted from a version used with elementary school teachers, with information on the performance of the items and scales when administered to a sample of preschool teachers. Depending on the outcome of the pilot, the measure may be used in future rounds of FACES. Teachers in the 120 programs that participated in the classroom-only data collection in spring 2105 were asked to complete a pilot version of the survey.

The survey measures five school organizational constructs that are referred to as the Five Essentials, and that are important for children’s learning and development:

- Inclusive leaders
- Building professional capacity through routines of collaboration
- Strong ties and partnerships among families, schools, and community
- Child-centered supportive environment
- Ambitious instruction
Tribal programs in Head Start Region XI have not been included in prior rounds of FACES. In recognition of the need for data on the children and families served by these programs, a first-ever study of Region XI Head Start was designed and conducted during the 2015-2016 program year.

AI/AN FACES is designed to inform policies and practices that address the needs of Region XI children and families. It will serve as a rich source of data about Region XI children and families. It is primarily descriptive, not meant as a study of the effectiveness of Region XI programs. Design and its implementation is a collaborative effort:
- Tribal program representatives
- University and Mathematica researchers
- Office of Head Start and OPRE staff

The design and implementation of the American Indian and Alaska Native Head Start Family and Child Experiences Survey or AI/AN FACES was a collaborative effort that involved the intense participation of tribal program representations, university and Mathematica researchers, and federal staff from the Office of Planning, Research and Evaluation and the Office of Head Start.
AI/AN FACES is a study of tribal programs in Head Start Region XI and the tribal and non-tribal children who attend those programs (86 percent of enrolled children are American Indian and Alaska Native). It is not a study of all American Indian or Alaska Native children who are served by Head Start. According to the 2014-2015 Head Start Program Information Report, 45 percent of the American Indian and Alaska Native children served by Head Start are enrolled in programs in Region XI, the remainder are enrolled in Region I-X programs.

A total of 1,123 children and their families were selected to participate in AI/AN FACES from 73 classrooms in 21 Region XI Head Start programs. Of these and among children whose parents consented to their participation in the study, 1,049 children and their families participated in AI/AN FACES. AI/AN FACES was conducted in fall and spring of the 2015–2016 Head Start year.

For the most part, the study used the instruments and protocols from the FACES 2014 Classroom + Child Outcome Core with some additions and modifications to capture key elements of tribal communities and programs and to ensure that the study is respectful of tribal customs and community members. Also, similar to the FACES 2014 Classroom + Child Outcomes design, children were assessed, their parents surveyed, and their teachers reported on their skills and behaviors in both fall and spring. Teachers and program and center directors answered surveys in the spring. Each child’s classroom was observed in the spring.

All data have been collected and data files will be archived at Research Connections.
Topic 1.3
FACES 2014 Sample
The design of FACES 2014 is comprised of two Core studies and one Plus study. (Here we focus only of the Plus study whose data are currently available through Research Connections.) Each of the studies answers different sets of questions that require different samples with some degree of overlap. The sample designs for the two Core studies and the Family Engagement Plus study are described in the following slides. More details on each of these samples can be found in the User’s Manual (See Chapter II, starts on pg. 41).
Each of the two Core studies on its own and in combination are nationally representative of Head Start programs, centers and classrooms, and of Head Start children for the Classroom + Child Outcomes Core. The 180 programs selected to participate in FACES 2014, including the 60 programs with child-level data collection and the 120 programs with no child-level data were selected from the final 2012-2013 Head Start Program Information Report (PIR) database.

A multi-stage sample design was used with four stages: (1) Head Start programs (grantees or delegate agencies providing direct services), (2) centers within programs, (3) classrooms within centers, and (4) children within classrooms (this final stage was only conducted in the 60 programs with child-level data collection). Programs and centers were selected using a probability proportional to size (PPS) approach and classrooms and children were selected with equal probability.

The samples for the two Core studies differed in several ways. The Classroom Core included larger program, center and classroom samples to support estimates at these levels. Children were only sampled in the 60 programs with child-level data collection (Classroom + Child Outcomes Core). Home visitors were eligible for sampling in the Classroom + Child Outcome Core, but not in the 120 programs that participated only in the Classroom Core study.

### Core Study Samples

- Nationally representative samples
- Use same program frame and same program eligibility criteria
- Use a multi-stage sampling approach
- The Classroom Core and Classroom + Child Outcomes Core samples differ in several ways
  - Number of programs, centers and classrooms sampled
  - Whether a sample of children is selected or not
  - Whether home visitors are eligible for sampling
This figure shows the flow of the sample selection for the two Core studies and the expected sample sizes at each level of sampling. Larger samples were initially selected to offset the impacts of non-participation. For example, 12 children were selected from each sampled classroom in order to have 10 children per classroom participating in the study.

Programs, centers and classrooms that were selected for the Classroom Core study will be revisited in the spring of 2017. Because some programs are no longer expected to be in operation at that time and programs will be providing services that were not doing so in spring 2015, the sample will be refreshed so that it is nationally representative of Head Start for program year 2016-2017. The data for this later round of the Classroom Core will be available for secondary analysis in summer 2018.
This table shows the sample sizes that were expected at each stage of sampling and the number that actually participated. Counts are provided for those sample units that were part of the Classroom + Child Outcomes Core (entered study in fall 2014) and those that only participated in the Classroom Core (entered study in spring 2015).

With a few exceptions, the number of participants met expectations. There were fewer classrooms and children than expected, which was due largely to the smaller program sample (n=176 vs. 180). The total number of centers and classrooms listed under the actual column (n=364 and 667, respectively) do not sum to the numbers in fall and spring, because a few programs/classrooms that were originally sampled and recruited became ineligible between the fall and spring.

**NOTE REGARDING THE LAST TWO ROWS OF THE TABLE:** The child sample was selected in fall 2014; no additional children were selected to participate in the study in spring 2015. However, not all children who were eligible and consented in the fall (n=2,462) were expected to continue to be a part of the study in spring 2015 (n=2,160). In reality, more children remained in the study in the spring (n=2,206) than expected.
This table shows the number and percent of study-eligible children whose parents consented to their participation in the study. In fall 2014, children were deemed eligible if they were enrolled in one of the FACES 2014 sampled centers when the child sample was selected. When children were sampled and recruited into the study, 93 percent of their parents consented to their participation by signing and returning the FACES 2014 parent consent form.

As in prior rounds of FACES, children who left the Head Start program or who moved to a Head Start program that was not in the FACES sample were not followed and were no longer eligible for the study. About 10 percent of the children with a parent consent left the program by spring 2015 (100 – (100*(2,206/2,462))). The remaining sample of 2,206 children represented 83 percent of the 2,645 children who were initially sampled for the study in fall 2014.
As the figures in this table show, the FACES 2014 sample is diverse. The sample sizes in the table do not sum to the total sample because of missing data. In addition, not all of the race/ethnic categories are listed here.

Primary home language is the language spoken most to the FACES child at home as reported by the child’s parent.

The counts of children in the sample with other characteristics can be found in the child codebook in Appendix G of the User’s Manual.
The Family Engagement Plus study was conducted in the 60 programs that were participating in the Classroom + Child Outcomes Core. A sample of 720 parents was selected for the qualitative interview. Initially one-half of the parent sample was released to be interviewed with the remainder serving as backups. Ultimately all 720 parents were released in order to come close to meeting the target of 360 completed interviews.

In addition, all parents were asked to complete one of two versions of a set of supplemental items as part of the spring 2015 Core parent and teacher survey, respectively. One-half of parents were randomly selected to receive each of the two versions, that are described in Module 2.

The target was 180 completed family service staff (FSS) interviews. A total of 196 family service staff were sampled and released in an attempt to meet this target. One half of FSS sample was randomly assigned to receive each of the two sets of Family Engagement interview modules, that are described in Module 2.
The Family Engagement Plus study was designed to provide more detailed information about parents’ engagement with Head Start and their service provision experiences. It also provides information about those staff who are the direct providers of services to parents and families.

While the sample sizes are adequate for these purposes, they are not large enough to support subgroup analyses of parents nor family service staff. The standard errors for subgroup estimates will be large and only very large differences will be statistically significant. Any analysis of subgroups should be considered exploratory and for hypothesis-generating purposes only.
Topic 1.4
FACES 2014 Data Collection
The two Core Studies – Classroom + Child Outcomes Core and Classroom Core use a variety of instruments to collect data on the children and families served by Head Start, and the programs and staff who serve them. They include:

- Direct assessments of children’s school readiness skills (language, literacy, math, executive functioning)
- Teacher reports on individual children’s social skills, problem behaviors, approaches to learning, academic accomplishments and special needs
- Parent surveys on children’s family, home, community, and Head Start experiences
- Head Start staff surveys (teachers and directors) on their background and experience, classroom or program characteristics, and policies and practices.
- Observations of Head Start classrooms to measure the quality children’s learning environment

All of these instruments are a part of the Classroom + Child Outcomes Core, but only the classroom observations and teacher and director surveys are administered to participants who are not part of the 60 programs with child-level data collection (Classroom Core).

More detailed information about the content of each of the instruments is included in Module 2.
### Data Collection Component Administration Characteristics: Core Studies

<table>
<thead>
<tr>
<th>Component</th>
<th>Type of Administration</th>
<th>Fall 2014</th>
<th>Spring 2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct child</td>
<td>Computer assisted</td>
<td>45 minutes</td>
<td>45 minutes</td>
</tr>
<tr>
<td>assessment</td>
<td>personal interview</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher child</td>
<td>Web with paper option</td>
<td>10 minutes per</td>
<td>10 minutes per</td>
</tr>
<tr>
<td>report</td>
<td>child</td>
<td></td>
<td>child</td>
</tr>
<tr>
<td>Parent survey</td>
<td>Web or computer assisted telephone interview</td>
<td>20 minutes</td>
<td>20 minutes</td>
</tr>
</tbody>
</table>

This slide and the next, identify the approaches used to administer the different Core study data collection instruments, list the amount of time that was required to complete each, and indicate whether the instrument was administered in both the fall and spring of the 2014-2015 program year or only in the spring. The three instruments listed in the table were only administered in the 60 programs with child-level data collection.

In both fall and spring, children in the Classroom + Child Outcomes Core participated in a 45 minute one-on-one direct child assessment. Child assessments were administered by trained field staff who completed a four and one-half day training prior to the fall data collection and a shorter refresher training in the spring. Assessors conducted the assessment by using a web-based instrument that they connected to using a tablet computer. Questions were displayed on their computer and a second monitor displayed images for the child (In the past, FACES used a paper easels with these images.).

Children’s primary classroom teacher used the Teacher Child Report (TCR) to provide information on each sampled child in their classroom in both the fall and spring. Teachers had the option of completing the TCR online or using a paper instrument. Sixty-six percent of TCRs were completed online in the fall and 84 percent were completed online in the spring.

Children’s parents also took a survey in both the fall and spring, which they could complete online or by telephone interview. The parent survey took on average 25 minutes to complete with 20 minutes devoted to Core study topics and the balance to topics introduced by the Family Engagement Plus study. In both the fall and spring about one-half of parents completed the survey online.

The content of each of these instruments and the measures used are described in Module 2.1 – FACES 2014 Core Study Instruments.
This slide identifies the approaches used to administer the four remaining Core study data collection instruments, lists the amount of time that was required to complete each, and indicates whether the instrument was administered in both the fall and spring of the 2014-2015 program year or only in the spring. The four instruments listed in the table were administered in all 176 programs that participated in FACES 2014. All four were administered in spring 2015 only.

Head Start teachers were asked to complete a teacher survey that they could do online or by using a paper questionnaire. Seventy-five percent of the teacher surveys were completed online.

Each of the sampled classrooms in all 176 programs was observed and the director of the program and its centers were surveyed. Each classroom was observed for about 4 hours by a trained classroom observer. Directors had the option of responding online or using paper questionnaires. Nearly all of the program directors chose the web option (97%) and 79 percent of center directors did the same.

Again, the content of each of these instruments and the measures used are described in Module 2.1 – FACES 2014 Core Study Instruments.
All Family Engagement Plus study data collection occurred in spring 2015 and included interviews with parents and family service staff (FSS). The parent and FSS interviews included both open-ended and closed-ended types of items. Data captured through the open-ended questions are not included on the FACES 2014 data files; only data from the close-ended items and constructed variables using these items are available on the data files.

In addition to the interviews, as part of the Core study spring data collection, the full sample of parents and teachers was asked to complete one of two sets of additional items that asked about their relationships and communication with teachers or about the types of social support they receive from family members and others and about the community/government services household members receive.
This slide identifies the approaches used to administer the different Family Engagement study data collection instruments, lists the amount of time that was required to complete each, and identifies how the data were collected. As noted earlier, all Family Engagement study data collection occurred in spring 2015.

Senior telephone interviewers at Mathematica’s Survey Operations Center conducted the parent interviews by phone using a paper guide. Family service staff interviews were conducted by members of the project team, again using a paper guide.

Supplemental items in the Core parent and teacher surveys were administered using the Core study procedures. Parents answered this set of items either online or by phone and teachers had the option of responding online or using a paper questionnaire.

### Data Collection Component Administration Characteristics: Family Engagement Study

<table>
<thead>
<tr>
<th>Component</th>
<th>Type of Administration</th>
<th>Spring 2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent Interview</td>
<td>Telephone interview with paper guide</td>
<td>60 minutes</td>
</tr>
<tr>
<td>Family Service Staff Interview</td>
<td>Telephone interview with paper guide</td>
<td>60 minutes</td>
</tr>
<tr>
<td>Parent survey items</td>
<td>Web or computer assisted telephone interview</td>
<td>5 minutes</td>
</tr>
<tr>
<td>Teacher survey items</td>
<td>Web with paper option</td>
<td>5 minutes</td>
</tr>
</tbody>
</table>
Module 1 Review

- To review some of the key themes and take-away-messages from Module 1, please answer the questions in the Module 1 Review Quiz. Reference slides at the end of this module.

- Once you have answered all the questions, check your answers with the Module 1 Review Quiz Answers. Reference slides at the end of this module.
Research Connections (www.researchconnections.org) provides users with extensive documentation on both the FACES dataset series (https://www.researchconnections.org/childcare/series/236) and FACES 2014. This documentation includes user’s manuals, questionnaires, and codebooks. These pages also provide information on accessing the datasets.

You will find more information about the purposes and history of FACES and its new design in the FACES 2014 User’s Manual. Here are the chapters where you will find this information. Also, each of the user’s manuals from earlier FACES cohorts contain background information about the study and its design. Other sources of information are the FACES technical reports and data tables that can be accessed at the Administration for Children and Families, FACES website (http://www.acf.hhs.gov/programs/opre/research/project/head-start-family-and-child-experiences-survey-faces) as well as on Research Connections (https://www.researchconnections.org/childcare/studies/36643). A bibliography of FACES publications is also available on Research Connections (https://www.researchconnections.org/childcare/resources/35542).

MODULE 1
REVIEW QUIZ & ANSWERS
Module 1 Review Quiz

1. FACES is a descriptive study of Head Start. It is not a study of Head Start’s impacts child and family outcomes.
   a. True
   b. False

2. Which of the following is true (Check all that apply).
   a. FACES includes national samples of Head Start Programs and children
   b. FACES data can be used to study changes in the children and families served by Head Start
   c. The design of FACES has not changed since it was first conducted in 1997
   d. Prior to 2014, the FACES sample represented all children served by the Head Start program

3. FACES 2014 introduced a new design that included (Check all that apply).
   a. Larger program and classroom samples
   b. An oversample of American Indian and Alaska Native children
   c. A greater focus on children’s Head Start experience
   d. Populations that had not been included in prior rounds of FACES
Module 1 Review Quiz Cont’d

4. Data from the Family Engagement parent and family service staff interviews allow researchers to test for differences between groups of parents and family service staff.
   a. True
   b. False

5. To answer questions about Head Start programs, centers and staff, researchers should use data from
   a. The 60 programs with child-level data
   b. The 116 programs with no child-level data
   c. All 176 programs with and without child-level
   d. None of the above, FACES 2014 is designed to study children and families, not Head Start programs, centers and staff
Module 1 Review Quiz Answers

1. ANSWER: True. FACES is not a study of the effectiveness of Head Start. The FACES sample is limited to Head Start children and their families. There is no control group or comparison group.

2. ANSWER: Only a and b are true. While relatively stable, the design of FACES has changed over the years. The biggest changes occurred with the introduction of the Core Plus study design that was introduced in FACES 2014. Therefore, c is not true. And, prior to 2014 the FACES sample only represented 3- and 4-year-old children who were entering Head Start for the first time; thus d is not true.

3. ANSWER: Response options a, c and d are true. FACES 2014 includes larger program and classroom samples (as well as a larger center sample). All data are now collected in a single year and there is no kindergarten follow-up as in the past; thus, a greater focus on Head Start. Family service staff have been included as part of the Family Engagement Plus study, a group that had not been included in prior rounds of FACES. There is no oversample of American Indian and Alaska Native children (b). AI/AN FACES is an independent sample of tribal programs, centers, classrooms/teachers, and children in Head Start Region XI. The Core study samples do not include Region XI and no special effort was made to increase the number of American Indian and Alaska Native children sampled from Region I-X programs.
Module 1 Review Quiz Answers, Cont’d

4  ANSWER: False. The sample sizes for these two interviews are not adequate to test for subgroup differences. Any such comparisons should be done only for exploratory purposes and to generate hypotheses for future research.

5  ANSWER: The correct answer is c (All 176 programs with and without child-level data). The Classroom Core sample, which includes all 176 programs, was designed to support program-, center-, and staff-level (program and center directors and Head Start teachers) analyses and to answer questions about these populations.
Head Start Family and Child Experiences Survey

FACES 2014

SELF-GUIDED TRAINING

Module 2.1

FACES 2014 Core Study Instruments

Prepared by
Jerry West, Ph.D.
University of Maryland, College Park

For Child Care and Early Education Research Connections,
National Center for Children in Poverty, Columbia University

With support from the Office of Planning, Research and Evaluation,
Administration for Children and Families,
U.S. Department of Health and Human Services

The development of these training materials benefited greatly from comments and suggestions provided by the staff of the Office of Planning, Research and Evaluation. Thanks is also extended to students at the graduate program of the College of Education, University of Maryland at College Park for their feedback and to participants in the summer data training program at the Inter-university Consortium for Political and Social Research, University of Michigan, who pilot tested the training. Any errors remain the responsibility of the Research Connections project.
This module introduces the FACES 2014 Core study instruments. It identifies the respondent for each and its content. The FACES child assessment battery that includes both direct and indirect (teacher reports) measures is described. Children follow different paths through the direct assessment depending on their home language, which is based on parent reports of the language spoken to the child in the home. The approach to deciding which path children follow is outlined. A more detailed description of the FACES child assessment battery can be found in the User’s Manual (See Chapter III, Section B, starting at pg.63).

Many of the items and measures used in FACES 2014 have been used in prior rounds of FACES. However, there were some revisions and additions to the child assessment battery. Also, the survey instruments include a few new topics and some topics that have been included in the past are receiving less attention this time. These changes and revisions are highlighted here.
The same six instruments listed here were also administered in previous rounds of FACES, although some of the questions and measures may differ from round to round.

The following sections highlight the topics covered by each of the Core study instruments. For each instrument, we identify the types of broad research questions that can be answered using the data that are collected by each instrument alone and in combination with data from other instruments.

The research questions, topics and types of data that are listed on the slides are not exhaustive. For a complete listing of the content of the different FACES 2014 instruments, review Appendix C (Instrument Content Matrices) of the User's Manual. The survey instruments can be found in Appendix D (Instruments) of the User’s. The child assessments and classroom observation measures are not included in the User’s Manual per agreements with instrument publishers and developers.
As described in Module 1, Topic 1.4, the direct child assessment, Teacher Child Report and parent survey were administered only in the 60 programs that were a part of the Classroom + Child Outcomes Core with data collected in fall 2014 and spring 2015. Assessments were conducted with all sampled children during visits to the 60 programs. Teachers completed a TCR for each FACES child in their class. Parents, whose children are participating in FACES were surveyed. The survey was completed by the parent who was most familiar with the child’s education and development. In a majority of the cases, this was the child’s biological mother. In some cases an adult household member such as the child’s grandparent completed the parent survey.

Lead classroom teachers, program directors and center directors in all 176 sampled and participating programs (programs with and without child-level data) completed a survey once, in the spring of the program year (spring 2015).

All sampled classrooms in the 176 programs were observed only in spring 2015. This includes classrooms that had FACES children enrolled and classrooms with no FACES children.

### Summary of Data Sources by Wave: Core Studies

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Fall 2014</th>
<th>Spring 2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct Child Assessment</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Teacher Child Report</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Parent Survey</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Teacher Survey</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Classroom Observation</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Program Director Survey</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Center Director Survey</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
Topic 2.1.1
Child Assessment Battery
One of the primary questions that drives the design of FACES centers on the school readiness skills of Head Start children when they first enter Head Start. And, what gains in these skills do they make while enrolled in Head Start? Because many of the measures chosen for FACES have population or developer norms, it is also possible to compare Head Start children’s skills and their growth in these skills to their peers of the same age.

FACES 2014, like earlier rounds of FACES, assesses children’s development in multiple domains, including language and literacy, mathematics, executive functions, socioemotional and physical. It uses a combination of direct and indirect assessments to measure children’s performance in these domains. Direct assessments measure children’s knowledge and skills based on their responses to questions (for example, pointing to a picture that matches a word spoken to them) and performance on specific tasks (for tapping a pencil/peg in response to the tapping of the person administering the assessment). Indirect assessments rely on others to describe and rate children’s skills. For example, teachers reporting on the social skills that they observe in a child during the program day.
Let’s first take a closer look at the direct child assessment used in FACES 2014.

Listed here is the full set of language and literacy assessments that are included in the assessment battery. They are used to assess children’s receptive (PPVT-4 and ROWPVT-4) and expressive (EOWPVT-4 and EOWPVT-4:SBE) vocabulary and early literacy (WJ III and Bateria III, Woodcock Munoz).

Items from the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) preschool wave that assess children’s knowledge of beginning and ending sounds, which were first included in the 2009 FACES assessment, are administered as a supplement to the WJ III Letter-Word Identification subtest. The items tap the skills of children who have progressed beyond letter knowledge but who have not yet acquired sight words. These items are only administered to a subset of children who respond correctly to 11 or more of the first 14 WJ III Letter-Word Identification items.

Detailed information about how children’s skills in each of these domains are measured and the psychometric properties of the measures used can be found in the User’s Manual (See Chapter III, Section B).
Children’s ability to analyze and solve practical problems in mathematics is assessed using either the Woodcock-Johnson III or the Bateria III, Woodcock Munoz: Applied Problems. Items from the preschool math assessment used in the ECLS-B are administered to capture math skills that are not tapped in the WJ III and Bateria III (for example, items that assess children’s understanding of relative size and ordinal numbers and their ability to recognize shapes).

FACES 2014 includes the pencil tapping task that was first introduced in FACES 2009. This measure of self-regulation, and in particular inhibitory control, is only administered to children who are age 4 or older at the time of the fall and spring assessments. Children who are 3 years old in the fall, but are 4 years old in the spring are administered pencil tapping in the spring only. The task requires the child do the opposite of what the assessor says (that is, tap one time when the assessor taps two times and tap two times when the assessor taps one time).

All children in the study are weighed and their height measured in both the fall and spring. These measures are used to calculate children’s height, weight and body mass index (BMI).

Detailed information about how children’s skills in each of these domains are measure and the psychometric properties of the measures used can be found in the User’s Manual (See Chapter III, Section B).

At the end of the one-on-one testing session, the assessor completes a set of rating scales from the Leiter-R to evaluate the child’s behavior in the test situation. FACES 2014 uses four of the eight subscales that make up the Leiter-R Cognitive/Social scale. The four subscales are listed on the slide.
The assessment begins with a language screener. The screener together with parent-reports of children’s home language is used to determine the pathway through the assessment that will give children the best opportunity to demonstrate their knowledge and skills. As in previous FACES, the language screener includes two subtests from the PreLAS 2000 (Simon Says and Art Show).

Children follow one of three paths through the direct child assessment (English, Spanish, non-English/non-Spanish). All children from a home where English is the primary language are routed to the English path. Children from a home where Spanish is the primary language are routed to the English path or to the Spanish path based on the number of errors that they make on the screener. Children who make more than 12 total errors across Simon Says and Art Show are routed to the Spanish Path, otherwise they are routed to the English path. Children from a home where a language other than English or Spanish is the primary language (for example, Chinese, Vietnamese, Croatian) are routed either to the English path or to a non-English/non-Spanish path if they make more than 12 total errors on the two screener subtests.

In the fall, both subtests are used for routing children to either an English, Spanish or Non-English/Non-Spanish path following the rules just described. In spring 2015, the full screener or both subtests are only used for children who did not pass the screener in the fall. In other words, children who previously demonstrated enough English proficiency to follow the English path through the assessment are not administered the full language screener in the spring. However, all children are administered Simon Says.

---

**Determining the Language of Assessment**

- Each child’s assessment begins with a language screener
  - Simon Says and Art Show from the PreLAS 2000
  - Used to determine whether child has sufficient English skills to be administered the assessments in English
- Children are routed to one of three different paths through the assessment
  - Based on child’s home language
  - Performance of the screener
- The approach used in the spring differs from that used in the fall
This table shows the assessment instruments that children receive based on their home language and their performance on the screener (i.e., the instruments that are administered to children who follow each of the three pathways through the assessment).

The PPVT is administered to all children, regardless of home language or performance on the screener. For children from Spanish-speaking households, the ROWPVT-4: SBE, a measure of their Spanish receptive vocabulary, is also administered, regardless of their performance on the language screener. After these two measures, the screener determines what language the child is assessed in. The language, literacy, math, and executive functioning measures all have a Spanish version or translation. ECLS-B Letter Sounds items are only administered to children on the English path and as noted earlier it is given based on performance on the WJ III Letter-Word Identification subtest.

Children on the other home language, non-English path (or non-English/non-Spanish path) receive an abbreviated assessment that only includes the two English language assessments (PPVT-4, EOWPVT-4) and height and weight measurements.

Height and weight are measured for all children.
This table shows the instruments that children from primarily Spanish-speaking homes are administered depending on whether they are routed to the English or Spanish path. Children who make more than 12 total errors across Simon Says and Art Show follow the Spanish path. Otherwise, they follow the English path. Both groups of children are administered the same three language measures (PPVT-4, ROWPVT-4:SBE, and EOWPVT-4: SBE). Both groups are administered the ECLS-B Math items and Pencil Tapping, and both have their height and weight measured. Children following the English path are administered the WJ III Letter Word, Spelling, and Applied Problems subtests and children following the Spanish path are administered the Spanish versions of these same subtests. The ECLS-B Letter Sounds items, which are available only in English, are only administered to the group of Spanish-speaking children who made fewer than 12 total errors on the screener and are following the English path.

The EOWPVT-4:SBE (test of expressive language) is conceptually scored which means children can provide responses in both English and their home language. Assessors can also provide prompts to children in both languages. If a child responds incorrectly in one language, they can be prompted for the correct response in their other language. Children with a Spanish home language may provide responses in both English and Spanish when administered the Bateria-III Problemas Aplicados and the ECLS-B math items. They may respond verbally or with fingers. Table III.4 in the User's Manual provides more information about the language administration and scoring approach used for each instrument in the FACES 2014 battery. See User’s Manual (Chapter III, Table III.4, pg. 70).

<table>
<thead>
<tr>
<th>English Path</th>
<th>Spanish Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPVT-4</td>
<td>PPVT-4</td>
</tr>
<tr>
<td>ROWPVT-4:SBE</td>
<td>ROWPVT-4:SBE</td>
</tr>
<tr>
<td>EOWPVT-4: SBE</td>
<td>EOWPVT-4: SBE</td>
</tr>
<tr>
<td>WJ III</td>
<td>Bateria-III</td>
</tr>
<tr>
<td>ECLS-B Letter Sounds</td>
<td></td>
</tr>
<tr>
<td>ECLS-B Math</td>
<td>ECLS-B Math</td>
</tr>
<tr>
<td>Pencil Tapping</td>
<td>Pencil Tapping</td>
</tr>
<tr>
<td>Height and Weight</td>
<td>Height and Weight</td>
</tr>
</tbody>
</table>

Spanish Children’s Pathway Through the Assessment
Most children in FACES are from homes in which English is the home language. However, a sizable number of children are from Spanish speaking households, and a small number are from homes where a language other than English or Spanish is spoken.

Many of the children from homes where a language other than English is spoken pass the English language screener and follow the English assessment path. The number of these children increases from fall to spring of the program year as children grow older and gain greater proficiency in English. For example, in fall 2014, about half of children from homes where a language other than English is spoken pass the language screener and are routed into the English version of the direct assessment. By the spring of 2015, far fewer are unable to be assessed in English (about 73 percent of these children are assessed in English in the spring).

This table shows the number of children who followed the three language pathways in fall 2014 and spring 2015. You should consider these results when analyzing the child assessment data and when interpreting the findings from your analyses. Children who pass the language screener and are routed into the English assessment path may differ from those who are not included in the English assessment.

### Language Routing Results

<table>
<thead>
<tr>
<th>Data Collection Wave</th>
<th>Home Language</th>
<th>Language Path</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>English</td>
<td>Spanish</td>
</tr>
<tr>
<td>Fall 2014</td>
<td>1,806</td>
<td>241</td>
</tr>
<tr>
<td>Spring 2015</td>
<td>1,593</td>
<td>338</td>
</tr>
</tbody>
</table>
FACES also includes indirect assessments of children’s social and emotional development and approaches to learning via individual Teacher Child Reports (TCRs) as well as questions about children’s accomplishments in the TCR. These items, which have been adapted from several existing measures (for example, the Social Skills Rating Scale, Personal Maturity Scale, Behavior Problems Index, and the Approaches to Learning Scale from the Early Childhood Longitudinal Study, Kindergarten Cohort), have been used in prior rounds of FACES.

The ratings of children’s accomplishments were adapted from items in the U.S. Department of Education, National Center for Education Statistics, National Household Education Survey. Detailed information about how children’s skills in each of these domains are measured and the psychometric properties of the measures are used can be found in the User’s Manual (See Chapter III, Section B, pg.80).
A goal of FACES is to provide accurate information about the population of children and families served by Head Start and how the population and its need for services may change. Maintaining consistency in the measures used to capture information about the Head Start population and its services, is key to achieving this goal. Thus, FACES 2014 uses many of the measures and follows many of the procedures that have been used in prior rounds of FACES.

However, several changes were made to the child assessment battery, which are listed on this slide. These changes were made to improve the accuracy of the data on children’s school readiness skills and in response to changes in the overall design of FACES. The Receptive One-Word Picture Vocabulary Test-4, Spanish-Bilingual Edition ROWPVT-4:SBE replaced the Test de Vocabulario en Imagenes Peabody (TVIP) replaced the Test de Vocabulario en Imagenes Peabody (TVIP) replaced the Test de Vocabulario en Imagenes Peabody (TVIP) replaced the Test de Vocabulario en Imagenes Peabody (TVIP) replaced the Test de Vocabulario en Imagenes Peabody (TVIP) replaced the Test de Vocabulario en Imagenes Peabody (TVIP). The fourth edition of the Expressive One-Word Picture Vocabulary Test is used, again providing more contemporary norms. The fourth edition of the Expressive One-Word Picture Vocabulary Test is used, again providing more contemporary norms.

Word Attack was dropped from the FACES battery. This measure was only used in the past when children left Head Start and entered kindergarten. Given the new design with all data collected during the Head Start year, the measure was no longer needed.

Parents no longer rate their child’s social skills, problem behaviors and accomplishments. FACES 2014 relies only on the ratings of children’s classroom teachers.

Finally, following an intensive study, the rule for determining whether children should following an English, Spanish, or non-English/non-Spanish path through the assessment was changed.

Chapter III of the User's Manual provides additional information on these changes and the rationale for each.
Topic 2.1.2
Core Parent Survey
The Core parent survey is the primary source of data on the demographic characteristics of children and their parents and children’s home learning environment. Parent respondents report on themselves and are proxy respondents for other members of the household, including the sampled child. Parents answer questions about their children that cannot be asked directly to the children because of their age. For example, their age, gender, and race/ethnicity.

Parent survey data alone and in combination with data from other sources are used to answer a wide range of questions that are of interest to the Administration for Children and Families and the Office of Head Start and to the early childhood community in general. The following slides include some of these questions and identify the types of data that can be found in the parent survey to answer them.
A key purpose of FACES is to describe the population of children and families who are served by Head Start, and if and how that is changing over time. The Core parent survey is the main source for information about the characteristics of Head Start children and their families.

Demographics of FACES children and their family members collected in the parent survey include race/ethnicity, nationality/country of origin, gender, and age. Parents are asked to complete a roster of all individuals living in the household, providing the number of adults and children in the home as well as their ages and relationships of adult household member to the sampled child. Parents are asked questions about the use of languages other than English in the home and what language is most often used when speaking and reading to the child. The parent survey also asks whether the children in the household attended Head Start or Early Head Start. Use of child care outside the Head Start day is gathered for center- and home-based (both non-relative and relative) care.
In addition to measuring children’s height and weight as a part of the direct child assessment, and to further understand children’s physical well-being, parents are asked about the overall health status of their child as well as their own mental health (using a short version of the Center for Epidemiologic Studies–Depression Scale).

**What is the status of children’s and families’ physical well-being?**

- **Child health and health care**
  - General health status
  - Who attends to child when sick and for routine medical and dental care
    - Regular health care provider
    - Health care and dental care providers
    - Where does child receive well-child care
- **Parent mental health**
The parent survey includes a number of questions that are designed to measure the different ways that parents support their children’s development. It includes items asking about the frequency with which parents and other family members read to their children and children’s access to books in the home.

The survey is also the source for information on a range of activities family members engage in with their children both in the home (for example, singing songs or making music and counting together) and outside of the home (for example, family outings, recreational and religious activities and shopping). Parents report on routines for certain activities such as meals and children’s bedtime and sleep quantity and quality.
The Core parent survey asks about total household income using a combination of open-ended and closed-ended questions. And, for the first time, FACES 2014 has included a measure of financial strain using items from the Economic Strain Questionnaire.

The survey also includes questions about the education and employment of the child’s mother and resident father. Questions are not asked about fathers who are not residing with the sampled child. Information is obtained on a family's food security.
In spring 2015, Head Start teachers, center and program directors are asked about the classrooms they teach or the centers and programs they manage. Each of these groups is also asked questions about their background, education and training and work experience. The following slides summarizes the types of questions and topics that are included in the Core teacher, center director, and program director surveys.

In a number of instances the same or similar questions or topics are included on both the center and program director surveys. This was done in order to get different perspectives on program policies and practices. For example, both program and center directors are asked about the length of time they spend on different responsibilities and on the ways that data are used in decision-making. However, the overall strategy was to ask the person who was most knowledgeable about a particular area. For example, center directors were asked questions that dealt with the ways teachers assess children and share results with parents and others. Program directors were asked about the curriculum and assessment tools used and the factors that were considered when selecting these. In addition, questions on a few high priority topics such as professional development are asked on all three staff surveys.
What are the characteristics and qualifications of Head Start Staff?

<table>
<thead>
<tr>
<th></th>
<th>Teacher</th>
<th>Center Director</th>
<th>Program Director</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics (gender, age, race-ethnicity, languages spoken)</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Employment and educational background</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Years of experience</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Salary and benefits</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Job satisfaction</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beliefs about teaching practice</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Depressive symptoms (CES-D short form)</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Children’s Head Start experiences likely are shaped by program staff, so it is important to understand the characteristics and qualifications of Head Start teachers and program and center directors. FACES obtains information on each through the teacher, program director, and center director surveys. Each survey contains questions about staff’s basic demographics, employment and educational background, years of experience, salary and benefits, and job satisfaction.

Teachers are asked about their beliefs about teaching practices and they answer the same set of questions about their feelings and moods that parents are asked.
Data from teachers, center and program directors can be used to examine differences in program management for programs of varying size. Center directors report on staffing, recruitment and turnover of lead and assistant teachers. They also answer questions about the staffing of bilingual teachers. Both center and program directors report on their responsibilities and the time required to carry these out.

Questions about participation in professional development are found in each survey as are questions about staff’s use of program data for decision making. Program directors provide information on the size of their program, including the total number of children who enrolled in the program and attended at least one class or participated in one home visit (cumulative program enrollment).
The teacher and director surveys ask questions about the main curriculum used and how children are assessed. Program directors are asked how the assessment tool was selected and teachers report on how the information from the assessment is used for planning classroom activities for Head Start children.

<table>
<thead>
<tr>
<th>What curricula and assessments are used in Head Start classrooms?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Name of curricula</td>
</tr>
<tr>
<td>Main child assessment tool</td>
</tr>
<tr>
<td>Methods used to assess &amp; frequency of assessment</td>
</tr>
<tr>
<td>Assessment used in planning</td>
</tr>
</tbody>
</table>
The successful implementation of curricula and assessment depends on teachers having the necessary knowledge and support. Training on curriculum is one item that is asked of Head Start teachers.

Professional development is captured from several sources and can be examined in isolation or across reporters. Professional development could also be examined to help explain variations in classroom quality.

Teachers and center directors are also asked whether they have received training pertaining to serving DLL children, which is a growing population in Head Start.
Topic 2.1.1
Classroom Observation
What is the quality of Head Start classrooms?

- Early Childhood Environment Rating Scale-Revised (ECERS-R), Short Version
  - Provisions for learning
  - Teaching and Interactions
- Classroom Assessment Scoring System (PreK CLASS)
- Classroom environment
  - Counts of children and adults
  - Literacy activities
  - Math activities
  - Instructional organization

FACES 2014 provides a national picture of the quality of Head Start classrooms. FACES classroom observation data may also be used to examine the relationship between classroom quality and children’s school readiness skills. Key to improving child outcomes is the quality of the classrooms in which they are served.

FACES 2014 used the main tools for observing quality in Head Start classrooms that were used in FACES 2009. It used a subset of 21 items from the ECERS–R, which is referred to as the ECERS–R, Short Version. These items form two factor scores: (1) provisions for learning, and (2) teaching and interactions. Chapter VII of the User’s Manual provides more information about these two scores and the items that were used to develop them. (See Chapter VII, Table VII.3, pgs. 232-234).

As in FACES 2009, FACES 2014 also used the CLASS, which taps three quality domains and six dimensions (in parentheses) of quality: (1) Instructional Support (concept development, language modeling), (2) Emotional Support (positive climate, teacher sensitivity) and (3) Organization (behavior management, instructional learning formats).

Observers counted the number of children and adults in the classroom at various times during their visit. The literacy and math activities in which teachers and children were engaged and the type of instruction that was being used (for example, whole or small group instruction, one-on-one instruction) was recorded.

FACES 2014 also has teacher reports on the time spent on specific literacy and math activities as part of the Core Teacher Survey.
FACES 2014 uses shorter and more streamlined surveys. As a consequence, some topics which had been included in past rounds of FACES are no longer included in the Core instruments and the depth of questioning around some topics has been reduced compared to the past. For example, questions about the services that programs provide families and the partnerships they form with other community organizations to supply these services are no longer included in the program director survey.

The parent survey still asks about household routines, but questions about child-rearing practices, parenting behaviors, and their involvement with the Head Start program are no longer asked. Some of these topics are now covered in the Family Engagement Plus study. For example, a subsample of parents answer questions about their families’ receipt of community services and sources of parent support. Questions about parents’ involvement with their child’s Head Start program and their satisfaction with Head Start are included in the Family Engagement parent interview.
While a number of topics that had been asked about in the past were not included in the Core surveys, some topics received greater attention than in the past. During the FACES Redesign, many stakeholder groups asked for more information about how programs were managed and funded and wanted to know more about program’s use of data in planning and decision making. As a result, items were added to the program and center director surveys on these topics. Stakeholders also asked for more information about teachers use of child assessments and the support they received in order to use these tools effectively.

Questions were added to the teacher survey about how child assessments were used and about the training and support teachers received on the child assessments that they administer to the children in their classrooms.
Module 2.1 Review

- To review some of the key themes and take-away-messages from Module 2.1, please answer the questions in the Module 2.1 Review Quiz. Reference slides at the end of this module.

- Once you have answered all the questions, check your answers using the Module 2.1 Review Quiz Answers. Reference slides at the end of this module.
Each of the Core study instruments is described in more detail in Chapter III of the User's Manual, starting at pg. 61. The content of each of the instruments is summarized in a set of matrices that you will find in Appendix C. The instruments themselves can be found in Appendix D. One way to identify the topics and questions found in prior rounds of FACES, but not in FACES 2014, is to compare the content matrices from FACES 2014 with those from FACES 2006 and 2009.

---

### Resources

- **FACES 2014 User’s Manual**
  - Chapter III – Data collection instruments
  - Appendix C – Instrument content matrices
  - Appendix D – Instruments
- **Instrument content matrices in the FACES 2006 and FACES 2009 user’s manuals**
- **Research Connections Resources**
MODULE 2.1
REVIEW QUIZ & ANSWERS
Module 2.1 Review Quiz

1. Several changes were made to the assessment battery for FACES 2014. Which of the following changes were made? (Check all that apply)
   a. The Receptive One-Word Picture Vocabulary Test-4, Spanish Bilingual Edition (ROWPVT-4, SBE) replaced the Test de Vocabulario en Imágenes Peabody (TVIP) as the measure of children’s Spanish receptive vocabulary
   b. Word Attack is no longer a part of the battery
   c. Teachers are no longer a source for information on children’s social skills and problem behaviors. FACES now relies on parent reports
   d. Only children who pass the language screener are routed to the Peabody Picture Vocabulary Test-4

2. Children who follow one path through the assessment in the fall (based on their home language and performance on the language screener) will follow the same path in the spring.
   a. True
   b. False

3. The designs of the Core studies with their emphasis on key indicator data and greater use of web surveys have led to longer parent, director and teacher surveys than in past rounds of FACES.
   a. True
   b. False
Module 2.1 Review Quiz, Cont’d

4. Which of the following are sources of information on professional development?
   a. Head Start teacher survey
   b. Parent survey
   c. Center director survey
   d. All of the above

5. If I am interested in the amount of time children spend on different reading and math activities in their Head Start classrooms, I should focus my attention on the Head Start teacher survey. It is the best source for such information.
   a. True
   b. False
Module 2.1 Review Quiz Answers

1  ANSWER: Only response options a and b describe actual changes that were made to the assessment battery. The ROWPVT-4 replaced the TVIP in order to get a more accurate picture of children’s Spanish language skills and more contemporary norms (a). Word Attack which was administered to children during the kindergarten round in the past is no longer needed now that there is no kindergarten follow up (b). Teachers in FACES 2014 are the sole reporter on children’s social skills and problem behaviors, not parents (c). All children regardless of their home language and score on the language screener are administered the PPVT-4 (d).

2  ANSWER: False. Children who follow the English path in the fall will follow the same path in the spring; however, children who did not follow this path in the fall will be evaluated again in the spring to determine whether they have the English language skills necessary to be administered the full battery in English.

3  ANSWER: False. The greater use of more streamlined, web surveys has meant that some topics that were included in the past have been dropped and the depth of questioning on other topics has been reduced. Some of these topics and questions are now covered in the Family Engagement Plus study.
Module 2.1 Review Quiz Answers Cont’d

4  ANSWER: Response options a and c identify sources for information on professional development. Questions on professional development activities are included in the Head Start teacher and center director surveys. These types of questions are not asked of parents. Another source is the program director survey.

5  ANSWER: True. While there are limited data on classroom activities collected as a part of the classroom observations (general reading and math activities children and teachers are engaged in during the observation period), teachers are asked about a wider range of activities related to reading, language and mathematics. They also report on whether the activities are scheduled and estimate the amount of time children spend on each.
Head Start Family and Child Experiences Survey

FACES 2014

SELF-GUIDED TRAINING

Module 2.2

FACES 2014 Family Engagement Plus Study Instruments

Prepared by

Jerry West, Ph.D.

University of Maryland, College Park

For Child Care and Early Education Research Connections,
National Center for Children in Poverty, Columbia University

With support from the Office of Planning, Research and Evaluation,
Administration for Children and Families,
U.S. Department of Health and Human Services

The development of these training materials benefited greatly from comments and suggestions provided by the staff of the Office of Planning, Research and Evaluation. Thanks is also extended to students at the graduate program of the College of Education, University of Maryland at College Park for their feedback and to participants in the summer data training program at the Inter-university Consortium for Political and Social Research, University of Michigan, who pilot tested the training. Any errors remain the responsibility of the Research Connections project.
The Family Engagement Plus study was designed to answer questions about the engagement and service provision experiences of Head Start families. This Plus study provides information about the efforts Head Start programs and staff make to support families and to engage parents in their children’s development. The study is designed to answer questions about the ways parents are engaged in their children’s learning inside and outside of the home. It also provides information about the direct providers of these services – family service staff.

The questions that can be answered, some of which are listed on the slide, require data that go beyond what are collected in the Core studies. However, in some cases, researchers will want to combine data from the Core study instruments (for example, parent survey or center director survey) with the data collected through the Family Engagement parent and family service staff interviews.

Family Engagement Research Questions

- What do family engagement efforts look like in Head Start?
- How are families engaged at home and in the community?
- What staff in programs are involved in family engagement efforts and in what ways?
- How are comprehensive services provided in Head Start?
- How do parents and staff characterize their relationships and interactions with one another?
Topic 2.2.1
Family Engagement Parent Interview
A 60–minute telephone interview was conducted with 315 parents who were sampled to participate in the Family Engagement Plus study. Key features of the interview and its administration are summarized on the slide. The interviews were conducted in English or Spanish. Remember, the parents selected to complete this interview (and their children) are not necessarily served by the family service staff participating in the study.
Parents were asked about their relationships and communication with family service staff (FSS), their Head Start experiences, and their perceptions of program practices. Open-ended questions from the Head Start Voices Pilot Study were used to solicit information about parents’ experiences with their child’s Head Start program and its services. Their relationships and communication with FSS were captured by close-ended survey items from the parent-family service staff version of the Family and Provider/Teacher Relationship Questionnaire (FPTRQ). The Family Engagement parent interview also included 16 items from the Strengths-Based Practices Inventory. Answers to these questions can be used to better understand the extent to which Head Start programs focus on family strengths and competencies rather than on deficits.

Key topics included in the interview are listed on this slide.

The Family Engagement parent interview did not ask demographic questions because these were asked in the Core parent survey and can be merged with data from the interview.
Topic 2.2.2
Family Service Staff Interview
A 60–minute telephone interview was conducted with 145 family service staff who were selected from the 60 programs participating in the Classroom + Child Outcomes Core. Key features of the interview and its administration are summarized on this slide. Again, as a reminder, the family service staff chosen for this interview are not necessarily serving the parents (and children) participating in the study.
Family service staff (FSS) were asked about their relationships and communications with families using items from the Voices Pilot Study (https://www.researchconnections.org/childcare/resources/27342) and the FSS version of the FPTRQ.

To understand more fully the association between the background and experience of FSS their relationships and communications with Head Start families, respondents were asked about their education, work experience, and languages spoken. The interview also collected basic demographic data.

This slide lists some of the topics found in the Family Engagement FSS interview. Knowledge (FSS knowledge about families served), Practices (FSS interaction and engagement with families), and Attitudes (FSS beliefs about families).
Topic 2.2.3
Parent and Teacher Core Surveys: Supplemental Items
Supplement items were added to the Core parent and teacher surveys in spring 2015. Parents answered questions about their relationships and communication with Head Start teachers, their family’s receipt of community services and different sources of social support. These items were combined to form two modules with one module being completed by each sampled parent in order to limit the burden on parent respondents.

Teachers were asked questions about their relationships and communications with parents. Unlike parents, all teachers responded to the same set of items.
The items in this module, which came from the parent-teacher version of the FPTRQ short form, focused on three areas: Knowledge (parents’ perspectives on teacher knowledge about the families served), Practices (parents’ perspectives on teacher interactions and engagement with families) and Attitudes (parents’ perspectives on teachers’ beliefs about families).
Parent Supplement: Social Support and Receipt of Community Services

- Parents answered questions about sources of social support when the family faces emotional, financial, and parent problems and emergency
- Parents were also asked about the community/government services household members received in the past 12 months such as
  - Job training
  - Child care
  - Transportation
  - Counseling
  - Medical and dental care

Items in this module asked parents where they turned for support when the family faces different types of problems and in emergencies. Parents were also asked about different services that household members received during the past year.
### Teacher Supplement: Relationships and Communication with Parents

- Head Start teachers reported on their relationships and communication with parents using items from the teacher version of the FPTRQ short form.
- Questions focused on three areas:
  - Knowledge
  - Practices
  - Attitudes

Teachers were asked a set of questions focused on three areas: Knowledge (teacher knowledge about the families served), Practices (teacher interaction and engagement with families) and Attitudes (teacher beliefs about families).
Module 2.2 Review

- To review some of the key themes and take-away-messages from Module 2.2, please answer the questions in the Module 2.2 Review Quiz. *Reference slides at the end of this module.*
- Once you have answered all the questions, check your answers using the Module 2.2 Review Quiz Answers. *Reference slides at the end of this module.*
Each of the Family Engagement study instruments (and the supplemental items included on the Core parent and teacher surveys) is described in more detail in User's Manual (See Chapter III, Section B, starting at pg. 103). The content of each of the instruments is summarized in a set of matrices that you will find in Appendix C. The instruments themselves can be found in Appendix D. The supplemental item sets are a part of the Core study instruments; they are not separate instruments in Appendix D.
MODULE 2.2
REVIEW QUIZ & ANSWERS
Module 2.2 Review Quiz

1. Which of the following questions could be explored using data from the Family Engagement Plus study (Check all that apply)?
   a. How are families engaged in Head Start and in their children’s learning?
   b. How are comprehensive services provided in Head Start?
   c. How do Head Start parents and staff characterize their relationships and interactions with one another?
   d. Do children and families whose services are provided by family service staff with more years of education and experience have more positive outcomes than those whose services are provided by less educated and less experienced staff?

2. In order to describe the background, education and work experience of parents who participated in the Family Engagement parent interview, data from the Core parent survey must be merged with data from the parent interview.
   a. True
   b. False

3. All parents and all teachers answered the same set of questions as part of the Family Engagement supplemental item sets.
   a. True
   b. False

   ANSWER: False. The items added to the parent and teacher surveys were not the same, and the supplemental items included in the parent survey were grouped into two modules with a random half of parents completing each of the two.
Module 2.2 Review Quiz Answers

1. ANSWER: Questions a, b and c could be examined using data from the Family Engagement parent and family service staff (FSS) interviews and supplemental questions on the Core parent and teacher survey. Question d requires that the sampled children and families be linked to their provider. This is not possible given the way that the FSS sample and Core child/parent sample were selected. The two samples are independent of one another. Children and families in the Core sample may or may not be in the caseloads of any of the FSS selected for the Plus study.

2. ANSWER: True. Background questions (e.g., age, race/ethnicity) and questions that ask about a parent’s education and employment were not included in the Family Engagement parent interview because they were asked in the Core parent survey. Because parents who were interviewed had also answered the Core survey, the data from both sources can be merged.

3. ANSWER: False. The items added to the parent and teacher surveys were not the same, and the supplemental items included in the parent survey were grouped into two modules with a random half of parents completing each of the two.
Head Start Family and Child Experiences Survey
FACES 2014
SELF-GUIDED TRAINING

Module 3

Data Access and Resources

Prepared by
Jerry West, Ph.D.
University of Maryland, College Park

For Child Care and Early Education Research Connections,
National Center for Children in Poverty, Columbia University

With support from the Office of Planning, Research and Evaluation,
Administration for Children and Families,
U.S. Department of Health and Human Services

The development of these training materials benefited greatly from comments and suggestions provided by the staff of the Office of Planning, Research and Evaluation. Thanks is also extended to students at the graduate program of the College of Education, University of Maryland at College Park for their feedback and to participants in the summer data training program at the Inter-university Consortium for Political and Social Research, University of Michigan, who pilot tested the training. Any errors remain the responsibility of the Research Connections project.
The overall goal of this module is to provide all data users, regardless of their level of expertise and experience, with an overview of the FACES 2014 data files and the contents of each. Data users will learn about each of the five FACES 2014 data files – three Core study and two Plus study files.

The module also introduces the constructed variables that are available on these files, such as classroom scores from the CLASS and ECERS-R, and the different types of scores derived from the direct and indirect child assessments. It identifies resources that are available to assist users in working with the data and describes the steps that are required to obtain access to FACES 2014 data from Research Connections.
The four topics included in Module 3 and a brief outline for each are listed on this slide.

See introduction for instructions on downloading the User’s Manual.
Topic 3.1
Data File Structure and Conventions
FACES 2014 data are organized into five separate data files. There are three Core study files, one for center/program data, one for classroom/teacher data and one for child data (including parent survey data). In addition, the data collected through the Family Engagement Plus study parent and family service staff interviews are stored on two separate files.
Each of the files contains a set of identifiers or ID variables that uniquely identify each case on the file (for example, each sampled child, each Head Start teacher and classroom). There is also a set of flags indicating whether there is a completed instrument for the case from each of the data sources.

The files contain item-level data from one or more of the study instruments. For example, the child file contains individual responses to the items on the fall and spring Core parent survey, and the center/program file includes responses to the items on the program director and center director surveys. Each file also includes a set of constructed/derived variables and scores that were developed from responses to questions asked in the appropriate surveys and interviews. For example, the classroom/teacher file includes all classroom/teacher constructed/derived variables (including classroom observation scores).

The child file is unique in that it includes not only assessment scores derived from the direct and indirect assessments, but also the constructed/derived variables from all Core data sources (parent surveys, teacher surveys, classroom observations, and center director surveys).

Each file contains the weight variables and the stratification and primary sampling unit (PSU) variables for the level of analysis required to calculate design-based standard errors using the data on the file. The number of weights on the files varies. Design-based standard errors and sampling weights are discussed in Module 4.

Finally, given that some of the instruments were only administered in 60 programs with child-level data (for example, parent surveys and child assessments), there are indicators in the program/center and classroom/teacher files that identify which programs have the child-level data and which do not.
This table summarizes the content of each of the three Core study data files. For each file, it shows the number of records (or cases), and identifies the source(s) of data found on the file. It also indicates whether the file includes both item-level data and constructed/derived variables using the source data (X), or only the constructed/derived variables (C). For example, the center/program contains item responses to the questions on the center and program director surveys, constructed/derived variables created from the center director survey and program-level aggregates of scores from classroom observations.

As noted earlier, the child file includes both item-level data and constructed/derived variables from all data sources. Because many of the measures used in the child assessments and classroom observations are copyright protected, only a limited amount of item-level data from these instruments is included.

The center/program file contains a record for each of the 347 centers that were eligible and sampled in fall 2014 or spring 2015. The classroom/teacher file contains a record for each of the 667 classrooms that were sampled and eligible for the spring 2015 data collection. This includes classrooms in the 60 programs with child-level data collection and classrooms in the 116 programs with no child-level data collection. The file contains another 24 classrooms that were not part of the classroom sample, but that children in the study moved to after sampling. The child file includes a record for each of the 2,462 eligible and consented children from the 60 programs with child-level data, regardless of whether there are data from the child assessment, TCR, or parent survey in fall 2014 or spring 2015.
The two Family Engagement files each contain a set of identifiers or ID variables that uniquely identify each case on the file (for example, sampled child, family service staff and Head Start program), sampling weights and item-level survey data from the appropriate interview. Each file also includes a set of constructed/derived variables from the Family Engagement interviews, which are described in a later slide.

The parent file has one record for each of the 313 parents who completed the Family Engagement interview as well as either the fall or spring Core parent survey. The FSS file has one record for each of the 145 completed FSS interviews.
The structure of each of the FACES 2014 files – Core study and Plus study files - is similar and begins with a set of identifier variables, followed by a set of flags indicating whether or not there is a completed instrument for the case from each relevant data source (for example, on the Core child file, there are flags indicating whether each child has a completed parent survey in fall 2014 and spring 2015), weight variables and the stratification and primary sampling unit (PSU) variables for the level of analysis required to calculate design-based standard errors, constructed/derived variables (when such variables were created), and individual item-level data from appropriate surveys or interviews.
As noted before, certain data that were collected as part of FACES 2014 are not included on the data files according to agreements with measure developers and publishers. FACES 2014 used several copyrighted instruments or scales (these include the direct child assessments, classroom observations, and indirect assessment of social skills and learning behaviors) and agreements with instrument developers and publishers sometimes did not allow individual items and item responses to be included on the data files. In these cases, only the scores derived from the individual responses are included. Those scores would be found within the “constructed/derived variables” portion of the data file: See User’s Manual for more information (Appendix J, pgs. 72-78).

Listed on this slide are the types of data that have been redacted from the data files that are available at Research Connections. Copyright protected items that have been redacted in the Teacher Child Report form are found in Appendix D of the User’s Manual.

Appendix D does not include the items administered to children in the direct child assessments nor does it include the FACES 2014 classroom observation protocol forms. However, chapters III and IV of the User’s Manual includes descriptions of the measures used in the child assessments and classroom observations and their administration. In addition, FACES uses mostly well known assessments and observation protocols, and information on these is widely available.
FACES 2014 follows a set of conventions when naming and labeling variables and assigning missing variable codes. Many were used in prior rounds of FACES. They are designed to aid researchers who are working with FACES 2014 data alone or in combination with data from earlier FACES cohorts.

The next set of slides provides more information on each of these conventions.
Most FACES 2014 variable names use a standard set of prefixes. The first character represents the source instrument such as the child assessment, parent survey or FSS interview. The second character in the prefix represents the wave in which the information was collected (1=Fall 2014 and 2=Spring 2015). You can use this to identify the source and data collection wave for each of the variables in the data files. For example, variable names beginning with P1 and O2 tells you that the data comes from the fall 2014 parent survey and the spring classroom observation, respectively.

For survey or interview items, the rest of the variable name identifies the questionnaire or interview item number in the source instrument. For example, P1D01 is based on item D1 in the fall parent survey which asks how often someone in the family reads to the child. In the case of constructed/derived variables, it can be a shortened description of the construct represented by the variable. For example, P1HHSIZE is number of people in the household and is derived from responses in the fall 2014 parent survey.

In general, all constructed/derived variables created from a given source use the same prefix as the source. Some such variables, however, combined information across data collection points and/or several sources and are not associated with any prefixes. For example, P1RMAGE for mother’s age uses information from the first completed parent survey (A majority of parent surveys were completed for the first time in fall 2014, but some were first answered in spring 2015). CHGENDER for child gender is created from information in the sample management system and the parent survey.

The FACES 2014 instruments contain many multipart questions. For example, the household information section in the parent survey asks the same set of questions for each household member. The conventions used to represent these multipart questions are described in chapter VI of the User's Manual (See Chapter IV, Section A.2, pg. 17).

A few exceptions that do not follow these conventions include identifiers (such as CHILLDID) or sampling weights (such as PRA1WT that is used when analyzing data from the Fall 2014 parent survey in combination with Teacher Child Report and child assessment data).
This slide lists the codes used as the first character in the FACES 2014 variable names. The source codes for FACES 2014 are the same as FACES 2009 and FACES 2006 with the exception of the Family Engagement interviews, which were not conducted before FACES 2014. Earlier FACES cohorts also used source codes in their variable names. The codes used in 2003 and before are listed in the last column of the table.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Source Code (1st digit)</th>
<th>FACES 2014, 2009 &amp; 2006</th>
<th>FACES 2003 or earlier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core child assessment</td>
<td>A</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>Core teacher child report</td>
<td>R</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>Core parent survey</td>
<td>P</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>Core Head Start teacher survey</td>
<td>T</td>
<td>L, F, H</td>
<td></td>
</tr>
<tr>
<td>Core classroom observation</td>
<td>O</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td>Core center director survey</td>
<td>C</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>Core program director survey</td>
<td>D</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>Family Engagement family services staff interview</td>
<td>F</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>Family Engagement parent interview</td>
<td>PE</td>
<td>--</td>
<td></td>
</tr>
</tbody>
</table>
This slide summarizes the conventions used when creating variable labels. The length of the labels is limited to ensure that they are compatible with most available software packages. As outlined in the slide, the approach used when creating labels for variables containing responses to questions/items in the survey and interview instruments is not the same as the approach used when creating labels for constructed/derived variables and scores.
There are three sources of missing data in FACES: unit nonresponse, item nonresponse, and legitimate skips. Unit nonresponse occurs when an entire instrument is missing for a case. For example, the parent of a sampled child did not complete a parent survey in the fall or there is no completed interview for a family services worker who was sampled for the Family Engagement study. Unit nonresponse in FACES 2014 data is addressed through the application of the sampling weights, which are discussed in Module 4.

Item nonresponse occurs when there are missing data on individual items that should have been responded to within a given instrument. Respondents may refuse to answer a question, state that they do not know the answer to the question, or they may simply skip the question. FACES uses three codes to distinguish these three reasons for missing item data (Refused, Don’t Know, and Not Ascertained).

With one exception, FACES does not impute for item level missing data. The one exception is household income that is collected in the parent survey. A data flag identifies those cases for whom household income has been imputed. The methods used to impute household income in the fall and spring are described in Chapter VII of the User's Manual (starts on pg. 209).

There are also items that are only answered based on a prior response or condition. For example, parents may be asked for more detailed information on their child’s ethnicity (such as whether child is Mexican, Puerto Rican or Cuban) if they first indicate that their child is of Hispanic or Latino origin. FACES assigns a different code for such data to distinguish them from other types of missing data.

You should determine how to handle data involving legitimate skips. Begin by reviewing the survey instruments to identify legitimate skips. How you handle these can affect your findings. For example, if you are interested in the percentage of all children whose parent usually speaks to them in Spanish at home, and you simply calculate the percentage from responses to question D10 in the parent survey, you will find that a large majority of children’s parents report speaking to them in Spanish. However, to answer this question correctly you must take into consideration that those parents who reported only speaking English in the home (Item D7 in the parent survey) skipped item D10. If we account for that legitimate skip and include those parents in the denominator, a much smaller percentage of children’s parents report Spanish as their first language.

How you handle missing data should always be aligned with your research question.
FACES 2014 data files are available in SPSS and SAS. Listed here are the established conventions for missing data codes in each file type. The codes listed in the table identify data that are missing because of legitimate skips, item nonresponse (missing data on items within a given instrument for one of three reasons (refusals, don’t know, and not ascertained) and unit nonresponse (an entire instrument is missing for the case). Cases are assigned a SYSMIS or “.” when an entire instrument is missing. For example, if a parent survey was completed in fall 2014, but a TCR was not completed for the same sampled child, all of the fall 2014 TCR data would be set to SYSMIS or “.” for that child.

<table>
<thead>
<tr>
<th>SPSS Code</th>
<th>SAS Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>.N</td>
<td>Not applicable, including legitimate skips</td>
</tr>
<tr>
<td>-7</td>
<td>.R</td>
<td>Refused (a type of item nonresponse)</td>
</tr>
<tr>
<td>-8</td>
<td>.D</td>
<td>Don’t Know (a type of item nonresponse)</td>
</tr>
<tr>
<td>-9</td>
<td>.M</td>
<td>Not Ascertained (a type of item nonresponse), items skipped but should have been answered</td>
</tr>
<tr>
<td>SYSMIS</td>
<td>.</td>
<td>System missing (unit nonresponse)</td>
</tr>
</tbody>
</table>
Each sampled program, center, classroom and child has a unique identifier. The lead teacher assigned to a classroom who is asked to complete the Head Start teacher survey and TCR also has a unique ID. The ID numbers are used in the data files to identify respondents for the different study instruments.

The child-level data file contains an identification number for each child and ID numbers for his/her Head Start classroom, teacher, center and program at the fall 2014 and spring 2015 data collection points. Because children may change classrooms and teachers between the fall and spring of the program year, they may have different values for the fall and spring classroom and teacher IDs, but in a majority of the cases, the values will be the same. The same identifiers are used on the separate data files for the classroom/teacher- and center/program-level data.

Module 4 includes a discussion of how to use the IDs to merge data from the different FACES 2014 files.
The Core parent survey includes a small set of questions that are only asked during the initial survey with the family. In most cases, the initial survey is completed in the fall, but for others, the first survey is completed in the spring. Included in this set are questions that collect demographic information on the child and the child’s mother and father, household income and family food security.

When constructing the Core child file, data on these items that were collected in the spring were merged with data collected in the fall. As a result, data on these items is contained in a single set of variables. The steps required to do this are outlined on the slide. For this limited set of variables, there are no fall and spring versions and the prefix for these variables has been changed to “Pn.” Flags are included with these variables indicating whether the data came from the fall or spring parent survey.

A description of this process can be found in the User’s Manual under “Data processing.” (See Chapter VI, Section 3, pgs. 174-175).
As described in chapter VII of the User's Manual, the data files contain a number of constructed/derived variables and assessment scores in addition to survey item-level data. This section introduces the constructed/derived variables and assessment scores that are found on the Core and Plus study data files. It reviews the reasons for using these variables and scores and identifies the types of variables and scores found on the files. Guidance is offered on which assessment score to use for different analyses and research questions.
There are an almost infinite number of variables and scores that could be created using the FACES 2014 data. The variables and scores on the data files were limited to those most critical for answering questions about Head Start children, families, classrooms, staff, and programs. Priority was given to variables or constructs that use data from multiple items or sources, and that require considerable effort to create. Some variables used a combination of data from survey instruments and from a proprietary survey management system (for example, children’s age and gender used data from this system and data from the Core parent survey).

Special statistical methods and software are needed to develop some of the scores found on the child file (for example, item response theory [IRT] analyses for assessment scores). Even with this knowledge and software, certain key variables could not be created by users because of restrictions on sharing individual item-level data from copyright-protected measures. For the most part, variables that are simple recodes or transformations of a survey item or question are not included in the constructed/derived variable set.

Many of the constructed/derived variables and assessment scores developed for FACES 2014 were used in prior rounds, while some are new to FACES 2014. Wherever possible, variables and scores used in prior rounds of FACES were created in the same way for this latest round of the study. This helps to facilitate comparisons of findings across rounds of FACES and when users include these variables and scores in their analyses it is easier to compare findings from different articles, papers or presentations.

### Why Use Constructed/Derived Variables and Assessment Scores

- Facilitate the use of FACES 2014 data in answering important questions about Head Start children, families, classrooms, staff and programs
- Eliminates the effort, and the special software sometimes needed, to create variables on your own
- Allows analysis of variables for which individual item data are not available on FACES 2014 data files
- Improves comparability of findings from one analysis to another and from one round of FACES to the next
The constructed/derived variables and scores created for the Core study fall into one of five groups, which are listed on this and the next slide. See User’s Manual for more information (Chapter VII, Table VII.1, starting on pg. 219).

Examples of constructed/derived variables on children and families include:

- Child and family characteristics such as variables that tap constructs such as the race/ethnicity, age, and gender of the sampled child and his/her mother, whether it was the child’s first or second year attending Head Start, mother and father education and employment, household income, family structure, and family economic risk.

- Family processes and parenting variables such as family-child activities, child care, health care access, parent depressive symptoms, food security, and household financial strain.

- The child file includes many different scores based on children’s performance on the direct child assessments, assessor and teacher ratings. Scores from the direct child assessments are measures of a child’s language, literacy, and math language skills, physical development and executive functioning.
Examples of constructed/derived variables about teachers, classrooms, and programs include:

- Teacher or classroom characteristics such as variables that come from one of two sources: (1) the Head Start teacher survey and (2) the Head Start classroom observation. The variables capture key aspects of Head Start classrooms and teachers, including class size, child/teacher and child/adult ratios, different scores depicting the quality of Head Start classrooms derived from the ECERS-R and CLASS, teacher depressive symptoms, and teacher beliefs. Several variables regarding the language environment of Head Start classrooms are included (for example, languages spoken by children, teachers and other adults, and language of instruction).

- Program characteristics such as derived variables focused on teacher turnover and teacher and family language use.
A number of constructed/derived variables and scores were developed from the Family Engagement parent interviews and supplemental items on the Core parent and teacher surveys. They measure important aspects of families’ engagement with Head Start and Head Start services, and the staff who work directly with families. The variables and scores are grouped into five categories, which are listed on this and the next slide, and can be found in the User’s Manual (See Chapter VII, Table VII.7, pgs. 259-265). The variables draw heavily on responses to questions from the parent-teacher and teacher versions of the Family Provider/Teacher Relationship Questionnaire, Short Form.

- Community services includes one variable that identifies the number of community or government services family members have received. This variable was derived from supplemental survey items on the spring Core parent survey.

- Parent social support includes several derived variables for the number of types of help parents receive from family members, friends and others, and whether parents find the support they receive helpful. These variables were derived from supplemental survey items on the spring Core parent survey.

- Parent-teacher relationships and communication includes a number of scale and subscale scores from responses to items on the FPTRQ, Short Form that was administered to a subset of parents and teachers as part of their Core surveys.
Strengths-based practices in programs includes empowerment, cultural competency, staff sensitivity-knowledge and relationship-supportiveness scores derived from responses to items from the Strengths-Based Practices Inventory in the Family Engagement study parent interview.

Parent-family service staff relationships and communications includes a series of scores derived from the FPTQR long- and short-forms that measure parents’ and family service staff’s perspectives on the relationships that exist between families and programs and the nature of the communications between the two. The data used in developing these scores come from the two Family Engagement interviews.

Due to an error in administration, eight FPTRQ items were not administered to FSS and one FPTRQ item was not administered to parents participating in the family engagement parent interviews. The omissions affect the ability to construct the following scores: FSS long-form Family-Specific Knowledge subscale scores, long-form Responsiveness subscale scores, and short-form Responsiveness subscale and Practices construct scores, as well as parent-FSS long-form Responsiveness subscale and Practices construct scores. Missing data for these items were imputed and FPTRQ FSS summary scores were created. A similar approach was not used for the one missing item in the family engagement parent interviews. More information on how these scores were created can be found in the User’s Manual (See Chapter VII, Section C, starting at pg. 254).
Assessment scores were created using children’s responses to the items on the direct child assessments of language, literacy and mathematics described in Module 2.1.

Child assessment scores in FACES include raw, standardized, and Item Response Theory scores (IRT-based scores) or W-scores. Which scores to use will depend on your research question and the type of analysis that you will be doing.

Raw scores refer to counts, averages, or the like of the individual items that a child completed. They are indicators of absolute rather than relative performance. In contrast, standardized scores allow for comparisons of an individual’s performance relative to others of the same age (or grade). There are two types of standardized scores on the Core child files: standard scores and T-scores. Standard scores have a mean of 100 and a standard deviation of 15. Scores above or below the mean indicate that compared to same-age peers, the child’s skills are more or less advanced, respectively. T-scores in FACES illustrate a child’s performance relative to the population of Head Start children as a whole, with a mean of 50 and a standard deviation of 10.

IRT scale scores (ECLS-B math assessment) and W scores (PPVT-4, Woodcock-Johnson III and Bateria III Letter-Word, Applied Problems and Spelling) are estimates of a child’s absolute performance.

There are two IRT-based scores of overall performance for the ECLS-B math items and letter-sound items. The IRT scale score is an estimate of the number of items a child would answer correctly if he/she had received all of the items in the ECLS-B preschool math assessment or all of the items in the Letter-Sounds section of the full ECLS-B battery. Math and letter-sound Theta scores, on the other hand, represent a child’s mathematics (letter-sound) ability based on the items that the child actually received. Generally speaking, Theta scores are more likely to be normally distributed than scale scores, making them more appropriate for certain analytic approaches, such as multi-variate regress and analysis of variance and factor analysis. However, scale scores are more easily interpreted and may be better suited for a broader audience.
In addition to the language, literacy and math scores, two sets of scores were created based on children’s responses to the Pencil Tapping task and using the measures of their height and weight.

There are two composite scores for executive functioning. One score is the number of times a child tapped correctly on the Pencil Tapping task. The second score is the percentage of times the child tapped correctly.

Children’s height and weight were measured twice (and in some instances, three times). The composite height and weight scores are the average of the two with some adjustments for measurements with larger discrepancies. They are expressed in inches (height) and pounds (weight). BMI was calculated using the composite height and weight measures and the Center for Disease Control and Prevention’s SAS program.

<table>
<thead>
<tr>
<th>Direct Assessment Scores: Executive Functioning and Physical Development</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Executive functioning (Pencil Tapping task)</strong></td>
</tr>
<tr>
<td>- Number of times a child tapped correctly</td>
</tr>
<tr>
<td>- Percentage of times the child tapped correctly</td>
</tr>
<tr>
<td><strong>Physical Development</strong></td>
</tr>
<tr>
<td>- Child’s height</td>
</tr>
<tr>
<td>- Child’s weight</td>
</tr>
<tr>
<td>- Child’s Body Mass Index (BMI)</td>
</tr>
</tbody>
</table>

Here is an example of how one would use the standard scores, which have a mean of 100 and a standard deviation of 15. This figure shows children’s scores on the vocabulary measures in fall 2014. On the left-hand side of the slide are children’s scores on the English receptive vocabulary measure (PPVT-4) and on the Spanish receptive vocabulary measure (ROWPVT-4: SBE). On the right-hand side are children’s scores on expressive vocabulary measures.

As was discussed in Module 2.1, all children in FACES are administered the PPVT-4, regardless of home language or performance on the language screener. In addition, children with a Spanish home language are administered the ROWPVT:SBE (receptive language) and the EOWPVT-4:SBE (expressive language), while those with an English or Other home language are administered the EOWPVT-4.

Looking first at the receptive vocabulary measures, we see that children are performing behind same-age peers (that is, below the mean of 100). They score approximately two-thirds of a standard deviation below national norms in the area of English receptive vocabulary (89.7). Meanwhile, children with a Spanish home language score more than one standard deviation below norms in the area of Spanish receptive vocabulary in the fall of the Head Start year (83.1).

In terms of their expressive vocabulary, children are performing closer to same-age peers. They score one-third of a standard deviation below norms on English expressive vocabulary (94.7) and those with a Spanish home language score near peers in the area of expressive vocabulary (97.5).
Here is another example, again using the standard scores. FACES captures information on children’s letter-word knowledge, early writing skills, and early math skills. The study uses the Woodcock Johnson measures with children who pass the screener and are assessed in English and the Woodcock Munoz measures with those who do not pass the language screener and are assessed in Spanish.

Regardless of the language of assessment, Head Start children’s literacy and math skills lag behind those of their same-age peers at the beginning of the program year. Children taking the assessment in English score at least one-third of a standard deviation below national norms on letter-word identification (93.8), early writing (90.3), and applied problems (93.1). Children taking the assessment in Spanish score two-thirds of a standard deviation below national norms on letter-word knowledge (96.7), and more than one standard deviation below norms on early writing (84.8) and applied problems (77.2).
A set of criterion or raw scores of children’s cognitive skills around emergent literacy and social-emotional development including social skills, problem behaviors, and approaches to learning were derived from responses on the Teacher Child Report. Scores capturing children’s social skills and problem behaviors are derived from several established rating scales for young children as discussed in Module 2.

Composite scores were calculated as the sum of the item responses (social skills and problem behaviors) of items and indicate the extent to which given statements reflect a child’s behavior. Similarly, teacher reports of children’s emergent literacy were added with sum scores providing a count of a child’s early literacy skills. Teacher-reported scores on approaches to learning were calculated as the mean of items and indicated the frequency with which given statements reflect a child’s behavior.

Assessor-reported scores of children’s behavior during the direct assessment included both raw and standard scores derived from the Leiter–R Examiner Rating Scale. The standard scores have a mean of 100 and a standard deviation of 15, indicating performance relative to same-age peers.
Use the Score or Scores that Best Answer Your Research Question

- IRT scale scores and to a lesser extent raw scores are useful for answering questions about the children’s skills have and gains in these over time
  - What skills do Head Start children demonstrate in the fall and spring of the program year?
  - What gains do they make in these skills over the program year?

- IRT scale scores are also used
  - To answer questions about how the skills of groups of children compare to one another at one point in time or over time
  - To explore the associations between multiple characteristics and one or more outcomes

With a variety of scores to choose from, especially scores derived from the direct language, literacy and math assessments, choosing the best score to use can be a little overwhelming. Each score offers a slightly different perspective on children’s development. The choice for the most appropriate score for analysis should be driven by the research question. Here are a few guidelines for choosing the score that is most suitable for your research question and analysis.

IRT scale scores are criterion referenced measures of absolute performance and are useful for answering the question, “What skills do Head Start children have?” IRT scale scores are also appropriate for studying gains in children’s skills over the program year. They can be used to answer the question, “How much have children learned or what gains do they make in particular domains from fall to spring of the year?”

In general, raw scores are not a very useful measure of absolute performance unless all children attempted to answer each of the questions in the assessment. Although they are perhaps the easiest to understand, they tend to underestimate performance.

IRT scale scores are also useful in identifying differences in the performance of groups of children at single points in time (for example, differences in the school readiness skills of children attending their first or second year of Head Start at the beginning of the program year). They can also be used to identify differences in the gains made by different groups of children. Finally, IRT scores provide a measure of performance that is useful in analysis of child outcomes that include multiple status variables, such as child and family demographics and family process variables.
Standardized scores (such as standard scores and T-scores) are overall measures of performance at a point in time, and are norm referenced. Unlike measures of absolute performance such as raw scores and IRT-based scores, they do not answer the question, “What skills do Head Start children have?” Instead, they answer the question, “How does the performance of Head Start children compare with the performance of their peers?” As such, standardized scores can be used to measure whether children are closing the achievement gap and whether some groups of children are making greater progress than others in closing the gap.
When your research question focuses on the development of children who are dual language learners and how their development compares to monolingual children you have to consider other factors when deciding on which assessment scores to use. For certain skills and assessments, you have scores for nearly all children (e.g., PPVT-4 and Pencil Tapping) because almost all children were administered the assessment regardless of home language or performance on the language screener. For other skills and assessments such as the WJ III Letter-Word Identification and Applied Problems assessments, this is not true. These assessments were administered to many fewer children whose primary language was not English, especially in the fall. While there were Spanish versions of these, the English and Spanish versions are not directly comparable and have different sets of norms.

The decision to administer the English or Spanish version of the assessments to a Spanish-speaking child was evaluated in the fall and again in the Spring. Children who were assessed in English in the fall were automatically assessed in English again in the spring. However, children who were assessed in Spanish in the fall could be assessed in either English or Spanish in the spring based on their performance on the spring language screener. Because the measures in the English and Spanish version of the assessment are not directly comparable, it is not possible to estimate changes in skill levels.

See Module 2.1 for more information on the language of assessment and the different language paths through the assessment.
Topic 3.3
Data User Resources
A number of resources are available to FACES data users. These resources, which are listed on this slide, are intended to provide researchers with the information they need to analyze the data from FACES 2014 accurately and to publish from this complex dataset. The following slides provide more information about each of these resources and their use.
The FACES 2014 User's Manual is the primary resource for information about the study design (Core study and Family Engagement Plus study) and its execution. It also includes useful information about the organization and structure of the data files and their use. The different sampling weights and how and when to use them are described. The data files include many scores that were developed from responses to the direct child assessments and the Teacher Child Report, and variables and scores derived from responses to the survey items and classroom observations. The manual describes how each was constructed and offers guidance on which assessment score to use for different research questions.

Regardless of your experiences using FACES data or data from large-scale studies such as FACES, you should begin by reading the “Getting Started” section of the User's Manual. It provides a brief introduction to the study and the new FACES design (Core Plus study design), contrasts FACES 2014 with earlier rounds of FACES, and directs the reader where to go to find more information on some of the major technical issues researchers will confront when working with these data.
The User's Manual also includes six appendices that provide additional information about FACES and the data on the five data files. The first four of these (Appendices A – B) focus on the study design, and in particular its instruments and measures.

Appendix A describes elements of the FACES design and its key measures. It provides this information for each of the rounds of FACES that were conducted from 1997 through 2014 and is an invaluable resource for users interested in questions that involve analyzing data across two or more cohorts.

Appendix B contains the copyright statements that should be included in any publication of the FACES 2014 data (and prior rounds as appropriate).

Appendix C contains an instrument content matrix for each of the study instruments and is an easy way to identify the types of measures and constructs that are included in FACES 2014. It also indicates whether a construct or measure was included in FACES 2009 or if it is new to FACES 2014. However, while information in the content matrices provides a way to quickly identify what kinds of questions might be answered using the FACES 2014 data, researchers should look at the actual items in the study instruments and be familiar with the different standardized instruments that are used in FACES (for example, PPVT-4, Woodcock-Johnson III, CLASS).

Appendix D includes a copy of each of the survey and interview instruments that was used in the Core and Plus studies. Copyrighted instruments or scales (to include the direct child assessments, classroom observations, and indirect assessment of social skills and behaviors) are NOT included per agreements with the instrument/scale developers and publishers. (See Chapter III, pg. 61 of the User’s Manual which details the original measures).
Appendices E-G contain a codebook for each of the three Core study files: Center/program, classroom/teacher and child. For each of the variables contained in these files, the codebook lists the variable name and label, the data format and type, and includes a set of descriptive statistics. See next section for examples from codebooks.

Appendices H and I contain a code book for the two Family Engagement Plus study files: Family service staff interview and parent interview. As a reminder, the supplemental items that were added to the Core parent and teacher surveys are included in the Core study files for those surveys.

Appendix J includes detailed information about each of the constructed/derived variables and assessment scores that are listed in the tables in Chapter VII of the User's Manual (Tables VII.1 – VII.7, starting at pg. 219).
EXAMPLE CODEBOOK
&
DESCRIPTIONS OF
CONSTRUCTED/DERIVED VARIABLES
### Example Codebook for Child-Level PUF

<table>
<thead>
<tr>
<th>Name</th>
<th>P1RDADED</th>
<th>Frequency</th>
<th>Code and Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label</td>
<td>P1: Father's Highest Education</td>
<td>381</td>
<td>1=Less than HS Diploma</td>
</tr>
<tr>
<td>Format</td>
<td>Num Type Construct</td>
<td>347</td>
<td>2=HS Diploma or GED</td>
</tr>
<tr>
<td>Valid N</td>
<td>990 Mean Maximum</td>
<td>205</td>
<td>3=Voc/Tech-Assoc-Some College Degree</td>
</tr>
<tr>
<td>Minimum</td>
<td>Maximum</td>
<td>57</td>
<td>4=Bachelor Degree or Higher</td>
</tr>
<tr>
<td></td>
<td></td>
<td>454</td>
<td>=System Missing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>969</td>
<td>-1/N=Not Applicable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>29</td>
<td>-9/M=Not Ascertained</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>P1RMAGE</th>
<th>Frequency</th>
<th>Code and Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label</td>
<td>P1: Mother's Age</td>
<td>2050</td>
<td>valid numeric value</td>
</tr>
<tr>
<td>Format</td>
<td>Num Type Construct</td>
<td>55</td>
<td>-9/M=Missing</td>
</tr>
<tr>
<td>Valid N</td>
<td>2000 Mean Maximum</td>
<td>357</td>
<td>SYSMIS/</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>P1RMOMED</th>
<th>Frequency</th>
<th>Code and Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label</td>
<td>P1: Mother's Highest Education</td>
<td>530</td>
<td>1=Less than HS Diploma</td>
</tr>
<tr>
<td>Format</td>
<td>Num Type Construct</td>
<td>633</td>
<td>2=HS Diploma or GED</td>
</tr>
<tr>
<td>Valid N</td>
<td>1931 Mean Maximum</td>
<td>617</td>
<td>3=Voc/Tech-Assoc-Some College Degree</td>
</tr>
<tr>
<td>Minimum</td>
<td>Maximum</td>
<td>151</td>
<td>4=Bachelor Degree or Higher</td>
</tr>
<tr>
<td></td>
<td></td>
<td>373</td>
<td>=System Missing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>144</td>
<td>-1/N=Not Applicable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14</td>
<td>-9/M=Not Ascertained</td>
</tr>
</tbody>
</table>
### Codebook for Child-Level PUF, continued

<table>
<thead>
<tr>
<th>Name</th>
<th>P1HHSIZE</th>
<th>Frequency</th>
<th>Code and Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label</td>
<td>P1: Household Size</td>
<td>1909</td>
<td>valid numeric value</td>
</tr>
<tr>
<td>Format</td>
<td>Num Type Construct</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid N</td>
<td>1909 Mean 4.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td>2 Maximum 12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>P1RFAGE</th>
<th>Frequency</th>
<th>Code and Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label</td>
<td>P1: Father's Age</td>
<td>1877</td>
<td>valid numeric value</td>
</tr>
<tr>
<td>Format</td>
<td>Num Type Construct</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid N</td>
<td>1877 Mean</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td>Maximum</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>P1DADEMP</th>
<th>Frequency</th>
<th>Code and Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label</td>
<td>P1: Father's Employment Status</td>
<td>566</td>
<td>1=Working Full Time</td>
</tr>
<tr>
<td>Format</td>
<td>Num Type Construct</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid N</td>
<td>889 Mean</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td>Maximum</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

566 1=Working Full Time
149 2=Working Part Time
75 3=Looking for Work
99 4=Not in Labor Force
553 .=System Missing
989 -1/..N=Not Applicable
31 -9/.M=Not Ascertained
## Appendix G, Descriptions of Constructed/Derived Variables

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Variable Label</th>
<th>Measure</th>
<th>ScoreType</th>
<th>Data Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>AnPPVcat</td>
<td>PPVT-4 Standard Score Distribution</td>
<td>PPVT-4</td>
<td>Categorical (integer)</td>
<td></td>
</tr>
</tbody>
</table>

### Direct child assessments

- **AnPPVcat**
  - **PPVT-4**
  - **Measure:** PPVT-4
  - **ScoreType:** Categorical (integer)
  - **Data Type:**
  
  **Description:** Standard scores allow for comparisons of an individual's performance to others of the same age (or grade). These scores have a mean of 100 and a standard deviation of 15. AnPPVcat categorizes children's standard scores using standard deviation units: scores less than or equal to 70 (at least two standard deviations below norms), 71 to 85 (between one and two standard deviations below norms), 86 to 100 (within one standard deviation of norms), and greater than or equal to 100 (at or above norms).

  **Specification:** Create a variable using the variable AnPPVT4S. Create categories 1) Standard Score LE 70, 2) Standard Score GT 70 AND LT 85, 3) Standard Score GE 85 AND LT 100, 4) Standard Score GE 100, and 5) NO BASAL.

  If AnPPV4NB = 1 then AnPPVcat = 5; else if AnPPV4S <= 70 then AnPPVcat = 1; else if 70 <= AnPPV4S < 85 then AnPPVcat = 2; else if 85 <= AnPPV4S < 100 then AnPPVcat = 3; else if AnPPV4S >= 100 then AnPPVcat = 4; else if AnPPV4S=M then AnPPVcat = M; else if AnPPV4NB NE 1 and AnPPV4S=N, then AnPPVcat = N else if AnPPV4S=***, then AnPPVcat = ***.
### Appendix G, Descriptions of Constructed/Derived Variables (cont’d)

<table>
<thead>
<tr>
<th>AnPPVT4R</th>
<th>PPVT-4 Raw Score</th>
<th>Continuous</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPVT-4</td>
<td>Raw Score</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>✔ Fall 2014 ✔ Spring 2015</td>
</tr>
</tbody>
</table>

**Description:** The Peabody Picture Vocabulary Test (PPVT-4) (Dunn, Dunn, and Dunn 2005) is designed to assess children's knowledge of the meaning of words by asking them to say or indicate, by pointing, which of four pictures best shows the meaning of a word that is said aloud by the assessor. PPVT raw scores are calculated as the last item in the highest set administered minus the number of errors. Raw scores can range from 0 to 225 and are an indicator of absolute rather than relative performance. This composite reflects child's raw score on the PPVT-4 assessment.

**Specification:**
- The ceiling rule for the PPVT-4 is that a ceiling is established if there are 8 or more errors in the item set, or the sum of correct items in the set is 4 or less. The Ceiling item number is identified as the last item in the highest ceiling item set.
- AnPPVT4R = the ceiling item (highest item administered) minus the number of errors.

For cases that do not establish a basal, AnPPVT4R or the PPVT-4 Raw Score is the sum of the items from the lowest item administered through the last item. These cases were flagged (AnPPVTNB).

For cases that were affected by an error in Blaise and did not reach ceiling (children ended assessment too early, should have been administered more items) in fall 2009, the PPVT-4 Raw Score is the sum of the items from the lowest item administered through the last item.

<table>
<thead>
<tr>
<th>AnPPVT4S</th>
<th>PPVT-4 Standard Score</th>
<th>Continuous</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPVT-4</td>
<td>Standard</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>✔ Fall 2014 ✔ Spring 2015</td>
</tr>
</tbody>
</table>

**Description:** The Peabody Picture Vocabulary Test (PPVT-4) (Dunn, Dunn, and Dunn 2005) is designed to assess children's knowledge of the meaning of words by asking them to say or indicate, by pointing, which of four pictures best shows the meaning of a word that is said aloud by the assessor. The PPVT was normed on a nationally representative sample of children and adults of various ages so that raw scores can be converted to age-adjusted, standardized scores with a mean of 100 and a standard deviation of 15. Standard scores can range from 20 to 160. PPVT standard scores indicate how an individual's score compares to the average score of people of the same age. This composite reflects children's standard score on the PPVT-4 assessment.

**Specification:** AnPPVT4S is constructed using a look-up table and is based on the child's age and raw score.
**Research Connections Resources**

- **Variable Search Tool**
  - View a list of all study variables
  - Search for particular variables of interest
  - View the accompanying variable label/question and response frequencies

- **“Resources Related to this Study”**
  - View a list of reports, papers, and other resources that cite the study
Topic 3.4
Access to FACES 2014 Data
Restricted-Use Data

- All FACES 2014 data files are restricted use
- Your application for access should include:
  1. Signed Restricted Data Use Agreement.
  2. A letter summarizing your research interest for the data requested
  3. Students must also include a photocopied student ID and co-signature of advisor or professor
- When approved, Research Connections will provide the data via secure download.
Module 3 Review

- To review some of the key themes and take-away messages from Module 3, please answer the questions in the Module 3 Review Quiz. Reference slides at the end of this module.

- Once you have answered all the questions, check your answers using the Module 3 Review Quiz Answers. Reference slides at the end of this module.
MODULE 3 REVIEW QUIZ & ANSWERS
Module 3 Review Quiz

1. Only one of the three Core study data files – the child file – includes the constructed/derived variables developed from all Core study instruments.
   a. True
   b. False

2. All Head Start teacher survey and classroom observation data are contained on the classroom/teacher file. The number of records on that file matches:
   a. the number of classrooms sampled in the 60 programs with child-level data
   b. the number of teachers in the FACES 2014 sample
   c. the number of classrooms selected from all 176 participating programs, plus classrooms that children moved to between the fall and spring of the Head Start year
   d. both the number of teachers and classrooms; the two are the same

3. The first two characters in the names of most of the FACES 2014 variables identify:
   a. the first two characters in the respondent’s name
   b. the data source for the variable and the wave in which the data were collected
   c. the program in the study and wave in which it first entered the study (fall 2014 or spring 2015)
   d. they have no specific meaning; they are just random characters
4. Which of the following are reasons for using the constructed/derived variables and scores provided to users on the data files?
   a. They can save you considerable time and effort
   b. They sometimes use data that are not available on the data file
   c. Using the variables and scores help when comparing findings from one FACES study to another
   d. All of the above

5. FACES includes several types of scores for the direct assessments of children’s language, literacy, and mathematics skills. Which of the following scores are measures of children’s absolute performance and which are measures of relative performance?
   a. Raw scores: ___Absolute ___Relative
   b. T-scores: ___Absolute ___Relative
   c. Standard scores: ___Absolute ___Relative
   d. IRT or W scores: ___Absolute ___Relative

6. You want to know if children have made progress over the program year and how much progress they have made on average. And, you want to know if the gains they have made are related to a set of family and classroom characteristics. What type of score is best suited to answer these questions?
   a. Raw score
   b. T-score
   c. Standard score
   d. IRT or W score
Module 3 Review Quiz Answers

1. ANSWER: True. All of the Core study files include constructed and derived variables, but only the child file includes both those created from the child assessments (direct and TRC) and parent survey, as well as any created from the teacher and director surveys and classroom observations.

2. ANSWER: The correct answer is c. The classroom/teacher file includes a record for each classroom that was sampled and a record for 24 classrooms that children moved into, which were not part of the classroom sample. When generating national estimates of Head Start classrooms and teachers, the 24 classes are not included (they do not have a positive sampling weight). Teacher data (for example, information on their background, education and work

3. ANSWER: The correct answer is b. The first character is the data source (for example, parent survey=P, child assessment=A, center director survey=C). The second character is the wave when the data were collected (fall=1 and spring=2).
Module 3 Review Quiz Answers, Cont’d

4  Answer: The correct answer is all of the above. Constructed/derived variables are timesavers (Some of the variables are complex and require the use of data from several sources). Certain data are simply not available on the data files (redacted data per agreements with test developers and publishers). With many different researchers using the data, measuring key constructs in the same way helps when comparing findings from one study to another.

5  Answer: Raw scores (a) and IRT or W scores (d) are both indicators of absolute performance and tell us what a child knows and how much he/she has learned over time. T-scores (b) and standard scores (c) are indicators of a child’s performance relative to others of his/her age. T-scores and standard scores tell us how a child is performing relative to their peers.

6  Answer: The correct answer is d. IRT-based scores (including W scores) are well suited for estimating the gains children make between fall and spring. They can be used as dependent variables in regression and other multivariate models. Raw scores could be used, but they are less meaningful unless all children attempt all of the items, which is not typically the case in the measures that are used in FACES.
Module 4

Sampling Weights and Variance Estimation

Prepared by
Jerry West, Ph.D.
University of Maryland, College Park

For Child Care and Early Education Research Connections,
National Center for Children in Poverty, Columbia University

With support from the Office of Planning, Research and Evaluation,
Administration for Children and Families,
U.S. Department of Health and Human Services

The development of these training materials benefited greatly from comments and suggestions provided by the staff of the Office of Planning, Research and Evaluation. Thanks is also extended to students at the graduate program of the College of Education, University of Maryland at College Park for their feedback and to participants in the summer data training program at the Inter-university Consortium for Political and Social Research, University of Michigan, who pilot tested the training. Any errors remain the responsibility of the Research Connections project.
The overall goal of this module is to assist researchers in developing accurate and reliable answers to questions about Head Start and the children and families who are served by the program. The module includes two topics that are central to this overall goal. The first topical module focuses on sampling weights. It begins with a general introduction to sampling weights with the goal of answering researchers’ questions about why they need to use sampling weights, and how they are created and used. This general introduction is followed by a description of the sampling weights included with the FACES 2014 data. It describes which of these weights to use when analyzing data at different levels and points in time, and when answering different types of research questions. Some of the most common questions users have about weights and their use are addressed.

The second topical module focuses on estimating variances and standard errors for complex sample designs, and is structured much like the module on sampling weights. First, researchers are introduced to the issues surrounding the estimation of variances and standard errors when the data come from a complex sample design, not a simple random sample (SRS). It introduces design-based methods for calculating standard errors and contrasts these with methods that assume the data come from a SRS. Second, it describes how design-based methods can be used to estimate standard errors for FACES 2014. Experienced researchers may want to opt out of those sections which introduce sampling weights and variance estimation in general. However, all researchers would benefit from reviewing those parts that deal directly with FACES 2014 sampling weights and variance estimation.
Topic 4.1
Sampling Weights
No topic generates more questions from secondary data users than sampling weights. The goal of this section is to provide researchers with a basic understanding of sampling weights and why they should use them when analyzing data from FACES 2014 and earlier rounds, as well as when analyzing data from other large-scale studies with complex sample designs. The material is organized around the questions and topics that are listed on the slide.
Sampling weights adjust for different features of complex sample designs like the one used in FACES and for nonresponse and sample attrition that is inherent in any national longitudinal study. Specifically, weights adjust for the fact that not all units had an equal chance of selection into the sample. They also adjust for differences in nonresponse among certain groups of the population and for sample attrition. They can help to reduce the potential for bias when there is differential nonresponse and/or attrition. In FACES, the basic sampling weight (inverse probability of selection) is adjusted for nonresponse at each stage of sampling and for instrument nonresponse (unit nonresponse).

Sampling weights adjust for unit nonresponse (an entire case or instrument is missing). They do not adjust for item nonresponse (individual items in an instrument have missing data). Different imputation methods are used to adjust for item-level missing data.

Weights are used when estimating characteristics of the population, which is the primary goal of FACES. FACES is not interested in the characteristics of the 176 programs, 667 classrooms, and 2,462 children in the study sample per se, but in the characteristics of the population of Head Start programs, classrooms and children. The data collected from sample cases when used with the appropriate sampling weights produce reliable estimates for the population of programs, centers, classrooms, teachers and other staff, and children.
Let’s continue our example, but now let’s include some data – Let’s say we want to estimate the average age of children in a class with 15 children. We could select a sample of the children and use their ages to estimate the average age of all 15 children in the class. If a simple random sample of the children in the classroom is selected, we could get an estimate of the average age of the children by calculating the simple mean of their ages. However, if a different sampling approach is used, one in which children might represent different numbers of children in the class or only themselves, calculating the average age without first adjusting for the differential probabilities of selection would result in an inaccurate estimate of the average age of 15 children in the class. As we will see later, using the weights also adjusts for other features of the survey such as nonresponse and sample attrition and can help to offset the potential bias associated with these.

This simple example shows the results of using the data from the same five children with and without weighting to estimate the age of the children in the class. The unweighted mean does not take into account the sample design where children were sampled with different probabilities and thus represent different numbers of children in the class. Here, each sampled child represents 1 to 5 children in the class. The weighted mean adjusts for these different probabilities of selection, and in our example using the weighted mean leads to a different and lower estimate of the average age of the class.

---

### Why are Weights Important?

- Using unweighted data may give you different and perhaps inaccurate results:

<table>
<thead>
<tr>
<th>Children’s age</th>
<th>4</th>
<th>3</th>
<th>3</th>
<th>5</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

- Unweighted mean: \((4+3+3+5+4)/5 = 3.8\)
- Weighted mean:
  
  \[
  \frac{(4\times1)+(3\times4)+(3\times5)+(5\times2)+(4\times3)}{15} = 3.5
  \]

---
As discussed in Module 1 (Topic 1.3 – FACES 2014 Sample), FACES is not a simple random sample of Head Start programs, classrooms and children. FACES uses a complex design that includes multi-levels of sampling and stratification to reduce the cost of the study and to ensure that the samples are representative of the population from which they were drawn. Sampling weights are used to compute estimates of the population (for example, average assessment scores, percentages of classrooms with a teacher who has a Bachelor’s degree, and average enrollments in Head Start programs) that reflect the sample design. So, what is a sampling weight and what is the difference between unweighted and weighted data?

When unweighted data are used, each observation is counted equally and the data represents only the observations of those programs and individuals in the sample. For example, the mean age of children attending Head Start in fall 2014 who are participating in FACES 2014 can be calculated by summing the ages of all of the children with data on the FACES child file and dividing by the number of children with age data. When weighted data are used, the observations are counted relative to sample members’ representation in the population from which the sample was drawn. Using the same data as before together with one of the FACES 2014 child-level weights we can estimate the mean age of the total population of children attending Head Start in fall 2014. We sum the weighted age of all children on the data file and divide by the sum of all the weights for those children.
Let's look at a simple example in order to examine further the differences between unweighted and weighted data.

If we want to know the mean age for all the children in one Head Start classroom, we could sum the ages of all the children in the classroom and divide by the number of children in the classroom. Alternatively, we could select a sample of the children and use their ages to estimate the average age of all the children in the class. Depending on how the sample is selected, some children might represent multiple children in the class and different numbers of children, and others might represent only themselves. Thus, calculating an unweighted (each child counts once) and weighted (different children represent different numbers of children in the class) averages would most likely produce a different result. NOTE: If we selected a simple random sample of the children in the class with each child having the same probability of being selected, the unweighted and weighted estimates would be the same.
There are several features of FACES and most large-scale national studies, that are important to keep in mind when considering to use weighted or unweighted data in your analysis.

- FACES is designed to produce national estimates for Head Start programs and centers, classrooms and teachers, and children and their families.
- FACES is a sample survey. It is not a census of Head Start programs, centers, classrooms and children. Only a small portion of each of these populations was selected and participated in the study.
- As mentioned several times before, FACES is not a simple random sample, and programs, classrooms and children did not have an equal probability of selection.
- Not all programs, staff, parents and children who were sampled or asked to complete a study instrument did so. There was unit nonresponse, which varied by instrument.
- Children who left their Head Start program after the fall data collection were not eligible for the spring follow up. Therefore, the sampling weights for the spring sample have been adjusted to reflect this.
The sampling weights developed for Core and Plus study data are designed for use in analyzing data at different levels and when using data from fall 2014 or spring 2015 alone (cross-sectional) or together (program year). There are also weights that support analyses of data from one or more of the Core study instruments and Family engagement interviews.

**FACES 2014 Weights Vary According to Several Factors**

- Level of analysis: program, center, classroom/teacher, program staff, child
- Waves of data collection: Fall or spring (cross-sectional) or both fall and spring (program year)
- Source(s) of data:
  - Core Study instruments – program/center director surveys, teacher surveys and teacher child reports, classroom observations, parent surveys, and child assessments
  - Family Engagement instruments – parent interview and family service staff interview
There are a total of 21 weights on the FACES 2014 data files, 17 Core study weights and 4 Family Engagement Plus study weights. A summary of the weights can be found in the User’s Manual (See Chapter VI, Sections B and C, pgs. 180-192). There are three sets of weights listed in the document: fall 2014 cross-sectional weights, spring 2015 cross-sectional weights and program year weights. The cross-sectional weights are used when analyzing data from a single wave of the study (fall 2014 or spring 2015) and the program year weights are used to analyze data across both waves. Core Study weights are included in each set and the Family Engagement study weights are only included in the spring 2015 and program year sets.

There are only three fall cross-sectional weights that support analyses of the fall Core child assessment, parent survey and Teacher Child Report (TCR) data. There are many more spring cross-sectional weights that support analyses using these same three Core instruments plus several additional instruments that were administered only in spring 2015 (for example, Head Start teacher and director surveys, classroom observations). The number of program year weights is again smaller, supporting longitudinal analyses of the Core child assessment, parent and TCR data.

The Family Engagement study spring cross-sectional weights support analysis of the family service staff interview data and analyses of the parent interview data together with Core parent survey data alone or in combination with data from the spring child assessment and TCRs. The single Family Engagement program year weight supports analyses using the data from both the Core parent survey and Family Engagement parent survey along with fall and spring child assessment or TCR data.

Descriptions of how these weights were developed can be found in the User’s Manual. (See Chapter VI, Section B, starting on pg. 180).
Deciding which of the 21 sampling weights to use can be a challenge, but there are a few simple rules that you can follow to make this task more manageable. First, and most important, the weight you choose should support the analysis that is required to answer your research question. More specifically, what is the target population and the level of analysis required to answer your research question(s)? What data sources will you be working with or where will you find data on the behaviors, outcomes, and characteristics of interest? Does answering your question require that you use one or two waves of data (is your focus on a single time point or on change between the fall and spring of the program year)? See User’s Manual (Chapter VI, Sections B and C, pgs. 180-192) to go back to the document you reviewed earlier. It describes the decision making process and is a handy reference to use when you are working with the FACES data.

Answering these questions will help you decide on which of the 21 weights is best for your purposes. You will not always find a weight that perfectly fits your needs. In many cases, you will need to decide between two or more weights, basing your decision on several factors such as how much missing data are acceptable and how important it is to have full information even if the number of cases must be reduced to do so. Comparing the number of cases that have a positive value for each of the weights you are considering will identify the impact of each weight on the size of your analytic sample. Examining the unweighted distributions for key variables for cases with a positive value on each of the different weights will quickly tell you how much missing data you will have if you use one weight versus another.

### Which Weight Should I Use?

- Weight should match your research question
- What is the level of analysis?
  - Program, center, classroom, teacher, staff, child/parent
- What data source(s) will be used?
  - See tables on the next two slides
- Does answering your question require one or two waves of data?
  - Fall 2014, spring 2015 or both
- There is not always a perfect weight!
As a reminder, data for the Core studies were collected in the fall and spring of the 2014-2015 program year (in the 60 programs with child-level data collection) or only in spring 2015 (in the 116 programs with no child-level data collection). Listed in the table are all of the Core study instruments along with information on when each was used.

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Fall 2014</th>
<th>Spring 2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct Child Assessment</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Teacher Child Report</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Parent Survey</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Teacher Survey</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Classroom Observation</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Program Director Survey</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Center Director Survey</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
There are fewer data sources for the Family Engagement Plus study and all data collection occurred in spring 2015.
Let’s take a look at several examples of research questions that can be answered with FACES 2014 and decide which weight is the best to use when answering each question.

The best weight to use for example 1 is P1_RA1WT. To answer the question you will be using data from a single wave (fall 2014). Furthermore, data on children’s language, literacy, and early math skills come from the direct child assessments. You are not using data from the Teacher Child Report, so there is no need to lose cases by selecting the weight that requires both the child assessment data and TCR data (PRA1WT). Also, the unit analysis here is the child, so you are opting for a child-level weight. Both cross-sectional child-level weights as well as all program year weights are conditioned on having a parent survey.
The best weight for example 2 is PRA12WT. The unit analysis here is the child, so you are opting for a child-level weight. You are not using data from the spring teacher survey or classroom observations. Therefore, PRA12OCW would not be the appropriate weight to use.
Example 3 – Choosing the Best Weight

- Are the gains children make in their language, literacy, and math skills during their time in Head Start influenced by the quality of their Head Start classrooms and teachers?
  - How many waves of data are needed?
  - What is/are the source(s) of these data?
  - What is the level of analysis?

- **Best weight: PRA12OCW**
  - To answer this question you will use data from the teacher survey and classroom observations together with data from the fall and spring direct child assessments
  - **PRA12OCW** is the only child-level program year weight that supports analyses using this combination of instruments

The best weight for example 3 is PRA12OCW. To answer the question you will be using data from both the fall and spring child assessment, and the spring classroom observations and teacher survey. The unit analysis here is the child, so you are opting for a program year, child-level weight.
Example 4 – Choosing the Best Weight

- Does classroom quality vary by classroom and teacher characteristics such as class size and teaching experience?
  - How many waves of data are needed?
  - What is/are the source(s) of these data?
  - What is the level of analysis?

- **Best weight:** TO2CLSWT
  - To answer the question you will use data from the spring classroom observations and teacher survey.
  - **TO2CLSWT is the only spring 2015 cross-sectional weight that supports analyses that use data from both instruments**

The best weight for example 4 is TO2CLSWT. To answer the question you will be using data from the spring classroom observations and teacher survey. All of these instruments were administered only in spring 2015, so you will be using a cross-sectional weight. The unit analysis here is the classroom, so you are opting for a classroom-level weight. In this analysis, teacher characteristics such as education, work experience and job satisfaction will be treated as classroom variables.

If your research question required that you only use data from the classroom observations or the teacher survey, you would use O2CLSWT or T2CLSWT, respectively.
Example 5 – Choosing the Best Weight

- In what ways are Head Start programs able to link data on family needs and services to child assessment information?
  - How many waves of data are needed?
  - What is/are the source(s) of these data?
  - What is the level of analysis?

- **Best weight: D2WT**
  - D2WT is the only weight that is provided for use in analyzing data from the program director survey.

The best weight to use here is D2WT, which is the only weight that is available for use when working with the program director survey data that are required to answer this question. The program director survey is administered only once, in the spring the program year (spring 2015). The unit analysis here is the program (or program director), so you are opting for a program-level weight.
The best weight to use here is a bit more complicated. To answer the question you could use data from several different sources, including the two Family Engagement interviews (parent and family service staff) and/or data from the supplemental items in the Core parent or teacher surveys. If you decide to focus on the responses to questions in the two interviews, you would use F2WT for the family service staff interview data and PE2WT for the parent interview data. PE2WT is also designed to be used when analyzing data from the Family Engagement parent interview together with data from the Core parent survey. You will need to use T2TCHWT if you want to use data from the supplemental items in the Core teacher survey. The data from each of these sources were collected only once, in the spring of the program year (spring 2015).

The three weights will support three different levels of analysis: child-level (PE2WT), staff-level (F2WT for family service staff), and teacher-level (T2TCHWT).

Choosing the best weight for your particular analysis can be challenging at times. You can make this task less difficult and stressful through practice. Reference slides at the end of this module for additional practice.
How to Use Weights with Commonly Used Software Packages

- **Use**
  - “WEIGHT” statement in SAS, or
  - “WEIGHT BY xxxxx” statement in SPSS
  - [weight=xxxxx] in Stata

- **FACES weights sum up to population totals**
  - Can be problematic for some statistical packages
  - Leads to incorrect standard errors and erroneous tests of significance

- **Normalizing weights**
  - Normalize weights for certain software packages (SPSS)
  - Normalized weights will sum to sample sizes, not population totals

FACES weights can be used with most commonly used statistical software packages. Shown on the slide are the weight statements that are found in SAS, SPSS, and Stata. In Stata, the typical statement is [weight] but there are 4 weight types (f=frequency weight; a=analysis weight; i=importance weight; p=probability weight) and the default varies by procedure. For frequencies the default is the “f” weight but it does not use non-integers values, so you will need to specify the “iweight” to get the weight applied accurately if it includes decimal points, which FACES weights do. For regressions, and correlation the default is “a” weight and it works fine with non-integer values. You should consult the software documentation for more information on how to incorporate weights in your analyses.

FACES weights sum to population totals, which can be quite large. Some statistical packages will use this weighted sum for the number of observations in the calculation of variances and standard errors. This can lead to incorrect and very small variances and standard errors and adversely affect tests of statistical significance. If the package you are using for your analyses uses the weighted n instead of the unweighted n or sample size, you will want to normalize your weights. The weight will now sum to the sample size and the unweighted n will be used when calculating variances and standard errors.
How Do I Normalize the Weight?

- Find the FACES weight that you want to use in your analysis
- Identify the number of cases with a positive value for this weight and the sum of the values of this weight
- Calculate a new weight using the following formula
  - \( \text{NewWeight} = \text{FACES weight} \times \frac{n}{\text{sum of FACES weight}} \)
- The sum of NewWeight will be the number of cases (sample size)

It is easy to normalize the weight. When you do this, you will be creating a new weight that sums to the sample size rather than to the population total. The new weight is equal to the FACES weight that you are using for your analysis multiplied by the ratio of the number of cases with this weight to the sum of the weight. You can use this weight when estimating means, percentages, and when doing correlational analyses, and the results for these estimates will be the same as when using the original weight. Standard errors for these estimates will now be calculated using the sum of the normalized weight (sample size).

A word of caution, the normalized weight will include non-integer values and some may be quite small. You should be sure that the procedure you are using for estimation purposes does not drop or round very small non-integer values.
Example - Normalizing Weights

- Weight to be normalized: PRA1WT
- Sum of PRA1WT weights = 769,686
- Total number of cases with a positive value for PRA1WT: 1,786
- Normalized weight =
  - NWPRA1WT = PRA1WT * (1,786 / 769,686)
- Sum of normalized weight = 1,786

Here is an example, using one of the sampling weights found on the FACES 2014 child file. PRA1WT is one of the three fall 2014 cross-sectional weights that is used to analyze data from the fall parent survey and data from the direct child assessment and/or Teacher Child Report. The values for this weight sum to the approximate number of children enrolled in Head Start in fall 2014. There are 1,786 cases in the data file with a positive value for this weight. Using these two values and the formula described in the previous slide a new weight (normalized weight) is created that now sums to 1,786 or the number of cases with a positive, non-zero PRA1WT weight.
Listed on this slide are some of the questions that researchers often ask about the use of sampling weights. Answers to these and other questions can be found in Sampling Weights: Answers to Frequently Asked Questions can be found in Slides 40-45. Many of these questions are also addressed in the FACES 2014 User’s Manual.
You should always use weights when analyzing any FACES data. This helps to ensure that your findings generalize to the populations represented by the different FACES samples (Head Start programs and centers, classrooms and teachers, family service staff, and children) and is one way that you can reduce the potential for nonresponse bias in your findings. You should use the weight that is appropriate for your research question, taking into account the level of analysis, waves of data, and the sources of the data you will be using.

FACES weights adjust for unit nonresponse (an entire instrument missing), but not for item nonresponse (items data missing for an otherwise completed instrument). For the most part, there are very little missing item data in the FACES. However, you should evaluate the amount of missing data on the variables you are using and decide how you want to address it in your analysis.

Various descriptive analyses, such as comparing the unweighted sample sizes for different weights and examining the unweighted distributions for key variables to identify the level of missing data, are useful tools when deciding between two or more weights. You may want to choose a weight that will maximize the number cases and the number of cases with complete data that are available for your analysis.
Topic 4.2
Variance Estimation
Data from FACES 2014 are used to produce population estimates for characteristics of Head Start, including its programs, centers, classrooms, teaching and non-teaching staff, and children. At each level of sampling (for example, program, classroom and child), the sample selected is only one of many that could have been chosen. Recognizing that estimates derived from a single sample may not represent the true value in the population from which the sample was drawn, researchers often compute the variance and standard error for the population estimates. These are used when placing confidence intervals around point estimates and when testing the statistical significance of group differences.

This section reviews reasons for estimating population variances and standard errors. It examines differences in standard errors for estimates derived from simple random sample versus complex sample designs. It describes design-based and approximation methods for estimating standard errors for complex sample designs like the one used in FACES.
FACES is a sample survey and the data collected from sample participants are used to estimate population characteristics (for example, average receptive and expressive vocabulary skills of Head Start children, percentage of Head Start teachers who have received training on their curriculum). Standard error is a measure of the variability or precision of those estimates. It indicates how close an estimate derived from one sample is to the actual value in the population. The smaller the standard error the more precise the estimate.

Researchers use FACES data to test a wide range of hypotheses about the characteristics of two or more groups of children, classrooms or programs, and the relationships between two or more variables. Standard errors are used when testing research hypotheses and when making inference to a population from the sample results.
Because standard errors play such an important role in research, it is important that they are accurate and that they capture the key features of the sample design. Many of the standard procedures found in commonly-used statistical packages assume that the data are from a simple random sample where the cases are assumed to be independent from one another and every case had an equal probability of being selected.

Classrooms and children in the FACES sample are not independent given the clustered nature of its design (for example, classrooms are clustered in program/centers and children are clustered in classrooms) and not all sampling units had the same probability of selection. Thus, standard procedures are not appropriate for estimating standard errors generated from the complex sample design used in FACES. Software packages designed for simple random samples tend to underestimate the standard errors for complex sample designs.

Using inaccurate standard errors when testing research hypotheses can lead to identifying statistically significant results where none are present (Type I Error) and vice versa (Type II Error). Using the correct standard error will help to reduce the chances of both types of error.
When estimating variances and standard errors for estimates derived from FACES, researchers should use a method that accounts for the FACES sample design. Over the years, FACES has used two design-based variance estimation methods. Prior to FACES 2006, a replication method was used and the data files included a set of replicate weights that researchers would use when estimating standard errors. Since then, FACES has used a Taylor Series methodology and data files include the design variables that are needed to calculate these weights.

It is strongly recommended that you use the Taylor Series method when estimating variances and standard errors for FACES 2014 (and for FACES 2006 and 2009 as well). However, if you do not have access to software that supports this method, you may use a method that approximates the impact of the design on these estimates. The method is simple to apply and the information you need to do so is included in Chapter VI of the User's Manual.
The table on this slide shows the standard errors for the percentage of children who are overweight and obese using the standard approach found in SAS (an approach that assumes simple random sampling) and the standard errors when the Taylor Series method is used. The table uses data from FACES 2009, but is included to illustrate the point.

As you can see, for children as a whole and for children in different racial/ethnic groups, the standard errors that are calculated using the standard method are smaller than those calculated with the Taylor Series method.
Section E (variance estimation) in Chapter VI of the FACES 2014 User’s Manual includes the specifications and code that you would use to calculate Taylor Series standard errors in SUDAAN and SAS. For each, it shows the weight, stratum and PSU variables that you would use for analyses conducted at the program, center, classroom, teacher and child levels. For analyses at the program level you need to specify the sampling weight (D2WT) and the first-stage sampling strata (STRAT). For analyses at the center, classroom or teacher levels, you need to specify the weight that is appropriate for your research question and both the strata and PSU variables (STRAT_C and PSU_C). For child-level analyses you need to use the appropriate weight and STRAT and PSU.

Special procedures in other software packages can also be used to generate accurate estimates for FACES 2014. Some of these are listed on this slide.
Whenever possible you should use the Taylor Series method to calculate standard errors for weighted estimates derived from FACES 2014 data. However, if that is not possible, you can use an approximation method. This approach is simple and straightforward, and involves adjusting the standard error generated by standard software (simple random sample standard error) by the design effect (DEFF), which is the ratio of the variance of a variable produced by specialized software that accounts for the complex design to the variance of the same dependent variable produced using procedures that assume simple random sampling.

If the software package you are using uses the weighted n (sum of the weights) rather than the unweighted n (sample size) when estimating variances and standard errors, you will need to normalize the weights and use this new weight to compute the standard errors before adjusting for DEFF. The normalized weight is created by multiplying the weight found on the FACES data file that you are using for your analysis by the ratio of the sample size (number of cases with a positive value on the FACES weight) to the sum of the values for the same FACES weight. The new weight (normalized weight) will sum to the sample size.

Caution! Certain procedures in available software (for example, proc univariate in SAS) will provide an erroneous number of cases if there are cases with a normalized weight equal that rounds to zero. Always check to confirm that the number of cases in your output is equal to the sum of the new weight.

When estimating the variance of an estimate you will adjust the variance by DEFF. When estimating the standard error you will adjust the standard error by the square root of the design effect or DEFT. Average overall FACES 2014 design effects (means and medians) for all levels of analysis and for select subgroups are provided in the FACES 2014 User’s Manual (See Chapter VI, Section F, pgs. 201-207).
This table shows the mean and median design effects that you would use when analyzing the spring 2015 Core study data. Other tables in the User's Manual include DEFF for different subgroups of children. There is no one correct DEFF to use when analyzing these data, but you should choose one that best aligns with the variables and subgroups in your analysis.
There are three ways that you can use DEFF or DEFT to adjust the simple random sample (SRS) variances and standard errors. You can multiple the SRS variance and standard error by DEFF and DEFT, respectively. The adjusted variance and standard errors would then be used when calculating the test statistic. Alternatively, you could adjust the t-statistic by dividing by DEFT or the F-statistic by dividing by DEFF. Lastly, you could adjust the sampling weight by dividing the sampling weight (or the normalized weight) by DEFF and then use this adjusted weight when calculating variances and standard errors.
Many software packages allow the use of sampling weights when doing regression analyses. However, when conducting regression analyses using FACES 2014 data, you want to use a software package that uses a Taylor Series method when calculating standard errors of the regression coefficients. Listed on this slide are several packages or routines within packages that do this. Again, if you do not have access to such a package, you could adjust the standard errors for packages that assume simple random sampling or adjust your sampling weight by DEFF and use this weight in our analyses. Once again, you may need to normalize your weights as well.
The sample design used to select the FACES sample is complex and is not a simple random sample. You should adjust for the complexities of the sample design when estimating standard errors by using the Taylor Series method or adjusting simple random sampling standard errors by the design effect (or the square root of the design effect or DEFT). You are strongly encouraged to use the Taylor Series method. Failing to adjust the standard errors by using one of these approaches can lead to inaccurate findings.

Some software packages use the weighted n (population count) rather than the unweighted n (sample size) when calculating standard errors (and variances). You need to be on the alert for this as it can lead to extremely small standard errors and incorrect test statistics.
Module 4 Review

- To review some of what we have learned about sampling weights and their use and about how to compute variances and standard errors for estimates generated from FACES 2014 data, please answer the questions in the Module 4 Review Quiz. Reference slides at the end of this module.

- Once you have answered all the questions, check your answers using Module 4 Review Quiz Answers. Reference slides at the end of this module.
Research Connections (www.researchconnections.org) provides users with extensive documentation on both the FACES dataset series: https://www.researchconnections.org/childcare/studies/36643. This documentation includes the User’s Manual, questionnaires, and codebooks. These pages also provide information on accessing the datasets.

The topics covered in this module are discussed in the FACES 2014 User’s Manual and in earlier FACES manuals. A detailed description of the complex design used to select the samples for the Core studies and Family Engagement Plus study is found in Chapter II of the Manual. Discussion of sampling weights and variance estimation is found in the following sections of the Manual:

• Sampling weights (See Chapter 2, Section B).
• Choosing the best weight (See Chapter 2, Section C).
• Variance estimation (See Chapter 2, Section E).

Appendix A contains a summary of key features of the FACES sample design and how it has changed and stayed the same over the years.

More details on the sample designs used in earlier rounds of FACES can be found in the FACES 1997 – 2009 user’s manuals.

FACES 1997-2009 data files also contain a set of sampling weights, which you can learn more about by reading the user’s manuals for those rounds of FACES. A different method was used to calculate standard errors in these earlier rounds of FACES (replication method). You can learn more about this method and its use by reading the manuals.
Frequently Asked Questions
Sampling Weights
1. When selecting a weight, do I have to subset my dataset?
ANSWER: There is no need to subset your dataset when using one of the weights on the FACES data files. Cases that are not assigned a valid weight (a non-zero, positive weight) have a missing value for the weight or a zero and would drop out of your analyses or contribute nothing to your estimates. That said, you could be cautious and select only those cases with a non-zero, positive weight.

2. What happens to cases where there is no positive weight?
ANSWER: Cases with a zero value or missing value for the weight variable do not contribute to the weighted estimates (for example, the case makes no contribution to either the numerator or denominator of a weighted average or weighted percent distribution). As noted above, a cautious approach is to select only those cases with a non-zero, positive weight when creating your analysis file.
3. What weights do I use if analyzing a subsample of cases?
   ANSWER: FACES weights were not designed for estimates of any particular subpopulation. Therefore, there is no guarantee that the weighted sum for the sample subpopulation will match that subpopulation in the FACES reference population (Head Start children enrolled in fall 2014), but the weights should still provide unbiased estimates of means, percentages, etc. for the subpopulation. You should select the weight that best matches your research question, following the same rules that were outlined in Module 4, Topic 4.1.

   One thing to keep in mind when analyzing a subpopulation (such as children enrolled in their first year of Head Start or children who stayed in the same classroom between fall and spring): it is best to read in the full sample file and specify in your statistical software the subpopulation to be analyzed (see bottom of page 164 of the FACES User’s Manual), rather than just keeping the cases you are interested in. That way, the complex sample design will be fully accounted for when calculating variances and standard errors.

4. What if I’m running a regression - what weights do I use?
   ANSWER: When running a regression analysis or any correlational analysis, you should select the weight that best aligns with your research question, following the same decision rules that were outlined in Module 4, Topic 4.1. You should choose your weight taking into account the rounds of data that you are using, the sources of the data, level of analysis and so forth.
FAQ

5. What weight do I use if I’m using a multi-level model?
ANSWER: Researchers will use multilevel modeling to examine the associations between Head Start classroom or program characteristics and child outcomes. Because FACES child weights take into account earlier stages of sampling (programs, centers, classrooms), you would only use a single child weight at level 1. That is, you do not need to use a child weight for level 1 and a different weight (for example, a classroom weight) at level 2. If you are doing a center- or classroom-level analysis, you would only need to weight the data with the appropriate center or classroom weight, respectively.

6. Does the use of imputed data affect how the weights are used?
ANSWER: Several factors must be considered when determining the weight to be applied when using imputed data. Were the missing data within an instrument (item-level missing) imputed and/or were all data missing because of the absence of a survey response (missing as a result of unit nonresponse)? The weights take care of “unit nonresponse,” meaning that we are missing an entire instrument, but do not take care of “item nonresponse,” meaning specific items being missing from an otherwise complete instrument. Some items are missing due to deliberate skip patterns, while others are missing due to refusals or “don’t know” responses, so you have to disaggregate the two. Unless you statistically impute values for the refusals and don’t know responses, cases missing one or more variables in your analysis will simply drop out of the estimate, potentially introducing bias. This is probably not a major concern for small amounts of missing data, but could be for larger amounts. If you have larger amounts of missing responses for a categorical response, you might want to create a missing value category with a unique designated (non-missing) value, and run your analysis that way.
FAQ

If the imputation was for an entire instrument (that originally was unit-missing), the use of a weight that includes that source instrument will likely result in the case still being dropped as the value of the weight for cases with unit missing is set to system missing. Here, you would want to use a weight that adjusts for the probability of selecting the case and perhaps for attrition in a longitudinal design, but that is not conditional on having a complete instrument for the data that are being imputed. If the imputation was done for individual items within an instrument (item-missingness), then that case would have a positive value for the weight. The use of weights in that scenario has no impact on your imputed data; it will weigh the case relative to its strength in the population based on its probability of selection.

7. For propensity score analysis, should weights be used at the point of creating the propensity score as the propensity score is sample specific?

ANSWER: At the time of creating a matched sample (either estimating the propensity score or using the propensity score to produce a matched sample), weights are not needed. The focus is to select the best match for each treatment unit for internal validity and matching as many treatment units as possible to have external validity. However, once you start analyzing the matched sample you would want to apply weights for generalizability to a larger population.
8. How are weights used in analysis using M-Plus or R?

ANSWER: M-Plus does allow the use of weights and the WEIGHTS option normalizes the weights automatically. For additional information for using weights related to variance estimation, please refer to http://www.statmodel.com/cgl-bin/discus/discus.cgi (In particular, the section on “multilevel data/complex samples”)

R can use weights if applied manually, but whether a given function in the package uses weights depends on the specific package you are using. Similarly, whether the particular command requires the analyst to normalize the weights (as compared to R doing it automatically) depends on the particular package. For variance calculations, design-based methods (replication, Taylor series) are supported by R, but again that depends on the package. The R survey package supports the Taylor Series method. Ultimately, the analyst should review the specific package documentation to determine if and how the functions within it handle weights.
Exercise
(with answers)

Which weight is the most appropriate?
Exercise Q&A

Instructions: Based on the research questions below, determine at what level the analysis will be conducted (for example, child, program, classroom), whether one (fall 2014 or spring 2015) or two (fall 2014 and spring 2015) waves of data are needed to support cross-sectional or program-year analysis, sample units that will be used (for example, program, classroom, teacher, child) and from what source(s) the data come (for example, child assessment, Teacher Child Report, family service staff). Using these pieces of information, determine which weight is the most appropriate.

1. Research question: When children enter Head Start, do their language, literacy, and early math skills differ by gender and race/ethnicity?
   Level of analysis: Child
   Wave(s) of data: Fall 2014
   Sample units: Children who were attending their first year of Head Start in fall 2014 (NEWTOHS=1)
   Source(s) of data: Child assessment, Parent survey
   Suggested weight: Fall 2014 cross-sectional weight P1_RA1WT. Because we don’t need TCR data, we can be less restrictive and use the weight that includes either TCR or child assessment data.

2. Research question: In the fall of the program year, do Head Start children’s social skills differ by home resources such as household income and number of books in the home?
   Level of analysis: Child
   Wave(s) of data: Fall 2014
   Sample units: Children who were attending their first or second year of Head Start (NEWTOHS=0,1)
   Source(s) of data: Parent Survey, Teacher Child Reports
   Suggested weight: Fall 2014 cross-sectional weight P1_RA1WT. Because we don’t need child assessment data, we can be less restrictive and use the weight that includes either TCR or child assessment data.
Exercise Q&A

3. Research question: Do the gains children make in their language, literacy, and math skills during their first or second year of Head Start differ by their skill level at program entry?
Level of analysis: Child
Wave(s) of data: Fall 2014, Spring 2015
Sample units: Children who were attending their first or second year of Head Start (NEWTOHS=0,1)
Source(s) of data: Child Assessment
Suggested weight: Program year weight PRA12WT. We are looking at gains across the first year or second year of Head Start, so we need a program year child-level weight. Because we only need child assessment data we can use a less restrictive weight that does not require teacher survey and observation data.

4. Research question: What is the relationship between children’s fall-spring gains in literacy and the quality of their Head Start classroom?
Level of analysis: Child
Wave(s) of data: Fall 2014, Spring 2015
Sample Units: Children who were attending their first or second year of Head Start (NEWTOHS=0,1) and whose classrooms were observed in Spring 2015
Source(s) of data: Child assessment, classroom observation
Suggested weight: Program year child-level weight PRA120CW.

5. Research question: What percent of Head Start teachers have a Bachelor’s degree? More than 3 years of experience teaching in a Head Start program?
Level of analysis: Teacher
Wave(s) of data: Spring 2015
Sample units: Head Start teachers
Source(s) of data: Teacher survey
Suggested weight: Spring 2015 teacher cross-sectional: T2TCHWT. We use this weight rather than T2CLSWT because the question is asking about teachers, not about the qualifications of teachers in Head Start classrooms.
6. How does Spanish-speaking dual language learners’ growth in English language skills compare to their growth in Spanish language skills during their second year of Head Start?
Level of analysis: Child
Wave(s) of data: Fall 2014 and Spring 2015
Sample units: Children who were attending their second year of Head Start (NEWTOHS=0)
Source(s) of data: Child assessment, Parent survey
Suggested weight: Program year weight PRA12WT.

7. What is the structural quality (e.g., child-to-teacher ratio) and process quality (e.g., instructional support) of Head Start classrooms?
Level of analysis: Class
Wave(s) of data: Spring 2015
Sample units: Head Start classrooms and their teachers
Source(s) of data: Classroom observation, teacher survey
Suggested weight: Spring cross-sectional weight TO2CLSWT. If we do not need teacher survey data, we can use the weight O2CLSWT.

8. What is the frequency of literacy and mathematics activities in Head Start classrooms?
Level of analysis: Class
Wave(s) of data: Spring 2015
Source(s) of data: Teacher survey
Sampling units: Head Start classrooms and their teachers
Suggested weight: Spring 2015 cross-sectional weight T2CLSWT. This weight is chosen because the data on classroom activities are collected in the teacher survey and because we are not using any classroom observation data.
9. What percentage of Head Start programs have a director with a Master’s Degree or above?
Level of analysis: Program
Wave(s) of data: Spring 2015
Source(s) of data: Program director survey
Sampling units: Head Start programs
Suggested weight: Spring 2015 cross-sectional weight D2WT. This is the only weight when using data from the
Core program director survey

10. What training have Head Start family service staff had to prepare them to identify the needs of the
children and families that they serve?
Level of analysis: Family Service Staff
Wave(s) of data: Spring 2015
Source(s) of data: Family service staff Interview
Sampling units: Family Service Staff
Suggested weight: Spring 2015 cross-sectional weight F2WT. This is the only weight when using data from the
family service staff interview alone.

11. How are families engaged in Head Start and in their children’s learning and development at home and in
the community?
Level of analysis: Child/parent
Wave(s) of data: Fall 2014 and/or Spring 2015
Source(s) of data: Parent survey, Family Engagement parent interview
Sampling units: Parents selected for the Family Engagement interview
Suggested weight: Program year weight PERA12WT. This is the most inclusive weight when the analysis includes
data collected from the Core parent survey and the Family Engagement parent interview.
MODULE 4
Review Quiz & Answers
Module 4 Review Quiz

1. As a rule of thumb, it is always best to use the FACES sampling weight that gives you the maximum number of cases for your analysis.
   a. True
   b. False

2. The sampling weights included with the FACES data adjust for the following (Check all that apply):
   a. Unequal probabilities of selection
   b. Parents unwillingness to consent to their child’s participation in the study
   c. Children leaving the study or leaving Head Start after the fall sample is selected
   d. Teachers failing to answer questions in the teacher survey about their age and salary

3. The weights included with the FACES data do not adjust for all possible combinations of missing data across the study instruments. You may need to choose a weight that is less than a perfect match for your research question.
   a. True
   b. False
Module 4 Review Quiz, Cont’d

4. Some software packages use the weighted n rather than the unweighted when calculating standard errors. What effects might this have on your findings? (Choose the correct answer)
   a. Standard errors would be much smaller than expected
   b. Standard errors would be small, but estimates of population means, percentages and correlations would be unaffected
   c. Both the standard errors and all other statistics (for example, population means, percentages, correlations) would be negatively impacted
   d. This has no real impact on the findings

5. It is always a good idea to adjust standard errors for design effects when analyzing FACES data even if the standard errors are estimated using a Taylor Series method.
   a. True
   b. False

6. When analyzing FACES data, it is best to use a design-based method to estimate the standard errors of means and percentages, but that is not necessary when running a regression analysis.
   a. True
   b. False
Module 4 Review Quiz Answers

1. **Answer:** False. You should choose the weight that is best suited for your research question, taking into account the unit of analysis, data sources and whether you will be using one or two rounds of data. When deciding between two weights that seem appropriate for answering your research question, you may want to consider how much missing data you are willing to accept, but this should not be your major reason for choosing the best weight.

2. **Answer:** The correct answers are a, b, and c. FACES weights adjust for different probabilities of being selected to participate in the study and for nonresponse among those selected. Parents' unwillingness to consent for their child to be in the study is one form of nonresponse (this affects the child consent or participation rate). The spring 2015 cross-sectional child weights and program year weights both adjust for sample attrition (children leaving the study) and for children leaving Head Start. Children who leave Head Start are no longer considered to be part of the population that is being studied.

3. **Answer:** True. Not all combinations of instruments and nonresponse are adjusted for in the weights that are included on the data files. There may not be a perfect weight for your specific analysis. When this is the case, you need to decide how much missing data you are willing to accept for specific instruments and variables in your analysis. It is often a good idea to compare the results from analyses using different weights when you have two to choose from and to examine the amount of missing data on your key variables.
Module 4 Review Quiz Answers Cont’d

4 Answer: The correct answer is b. Standard errors would be underestimated and test statistics such as t-values would be inflated, but this has no effect on the estimates for other statistics (means, percentages, correlations, regression coefficients). The latter would be estimated correctly.

5 Answer: False. The Taylor Series method adjusts for the complex sample design used by FACES. There is no reason to make any future adjustments to these standard errors.

6 Answer: False. Design-based methods should be used to estimate the standard errors of regression coefficients. If your software does not support Taylor series, the standard errors should be adjusted by the design effort using one of the three approximation methods discussed in Module 4.
Tips for Working with the Data

Prepared by
Jerry West, Ph.D.
University of Maryland, College Park

For Child Care and Early Education Research Connections,
National Center for Children in Poverty, Columbia University

With support from the Office of Planning, Research and Evaluation,
Administration for Children and Families,
U.S. Department of Health and Human Services

The development of these training materials benefited greatly from comments and suggestions provided by the staff of the Office of Planning, Research and Evaluation. Thanks is also extended to students at the graduate program of the College of Education, University of Maryland at College Park for their feedback and to participants in the summer data training program at the Inter-university Consortium for Political and Social Research, University of Michigan, who pilot tested the training. Any errors remain the responsibility of the Research Connections project.
The objectives of this module are to introduce some of the technical issues you will need to be familiar with when using FACES data and to assist you in producing accurate and reliable answers to their research questions. A set of guiding principles are offered that should lead to high quality analysis of FACES data.
This module includes the four topics listed on the slide. They are designed to assist data users who need to merge data from the different FACES 2014 files in order to answer their research question and those whose question requires that they conduct teacher-level and/or class-level analyses of data captured by the Head Start teacher survey and through classroom observation. It is also directed towards researchers who are interested in analyzing data from FACES 2014 together with data from earlier rounds of FACES (e.g., FACES 2000 through FACES 2009) to study changes or trends in Head Start over time.

The last topic covered in Module 5 offers researchers suggestions for how they should approach working with FACES 2014 data in order to answer their research questions accurately. The material covered in each of these topical modules is appropriate for researchers with all levels of experience and expertise.
Module 3 described the contents of each of the five FACES 2014 data files. This section begins with a brief review of the contents of each file and then describes how data from the different files may be merged when your research question requires the use of data that are not contained on a single file.
This table summarizes the content of each of the three Core study data files. It identifies the source(s) of data found on each file. It also indicates whether the file includes item-level data (responses to questions on the different instruments and the constructed/derived variables created from the source data (X), or only the constructed/derived variables (C).

For example, the center/program file contains item responses to the questions on the center and program director surveys, constructed/derived variables created from the center director survey and program-level aggregates of scores from classroom observations. The child file includes both item-level data (primarily from the parent survey) and constructed/derived variables from all Core data sources. It does not include item-level data from these other sources (Head Start teacher survey, program and center director surveys).
The two Family Engagement files each contains item-level data for closed-ended questions (source data) from the appropriate interview and constructed/derived variables from responses to the closed-ended items in the interview (for example, scale and subscale scores from the FPTRQ).

Responses to the supplemental items that were administered as part of the spring Core parent and teacher surveys are not included on these files. They are found on the Core files for the two surveys.
The FACES data files were created to minimize the number of files that researchers would be required to work with in order to answer their research questions. For example, because many FACES data users will be interested in the effects of different environments (classroom, home, program) on children’s development, the child-level file includes all of the parent interview data and the constructed/derived variables created from all of the other Core data sources. Nevertheless, because there are an infinite number of research questions and many variables stored across the five data files that might be used to answer these questions, it is inevitable that some questions will require researchers to merge data from one or more of the data files.

Listed on the slide are two reasons why researchers will need to merge data from one file with data from another. First, many source items are only included on a single file. So, for example, researchers who want to examine the association between the domains and dimensions of classroom quality and program characteristics (for example program size and program professional development support) will need to merge data from the classroom/teacher and center/program files.

Second, some research questions may require that data collected at one level be aggregated and analyzed at a higher level. For example, researchers may be interested in the support provided by programs for professional development and whether it varies by the proportion of a program’s teachers with a B.A. degree. To answer this question, data on teachers’ education from the classroom/teacher will need to be aggregated to the program level and merged with data on program support for professional development that is stored on the center/program file.
To merge data across files and to aggregate data collected at lower levels, you will need to use the identifiers provided on the data files. As was noted in Module 3, the child data file contains an identification number for each child and ID numbers for the Head Start program and center they attended in spring 2015, and their classroom and teacher in both fall 2014 and spring 2015. The same identifiers are used on the separate data files with the exception of the classroom identifier in the classroom/teacher file. The classroom identifier on that file is CLS_ID and data from the child file may be merged with classroom data using the fall and spring identifiers (CLS1_ID and CLS2_ID).

Section A.6 (Merging data from different files) in Chapter VI of the FACES 2014 User’s Manual includes detailed information on merging data from the different files.

### Merging Files: Identifiers

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Core Study File</th>
<th>Family Engagement File</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Center/Program</td>
<td>Classroom/Teacher</td>
</tr>
<tr>
<td>D2_ID</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>C2_ID</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>CLS1_ID</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>CLS2_ID</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLS_ID</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1_ID</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>T2_ID</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ChildID</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Topic 5.2
Teacher- vs. Classroom-Level Analysis
Data from the spring 2015 Head Start Teacher Survey can be used to answer a wide range of questions about the population of teachers working in Head Start programs and the classrooms that they teach. Because some teachers in the study taught more than one sampled classroom, there are data on more classrooms than teachers. This section describes the steps that are necessary when analyzing the data on the classroom/teacher file to answer questions about Head Start teachers versus Head Start classrooms.

The lead teacher in each sampled classroom completed a questionnaire that included questions about their classroom and themselves. Some teachers were teaching two half-day classes in spring 2015 (n = 23). This group of teachers answered questions about each of their classes (i.e., AM and PM). All teachers provided information about their background/demographics and experience only once. Teacher-level weights adjust for duplication of classroom information for these teachers.

The classroom/teacher file contains another 24 classrooms that were not part of the classroom sample, but that children in the study moved to after sampling. There are no classroom or teacher weights for these classrooms (and their teachers). Thus, they would not be included in any analyses that are conducted at the teacher- or classroom-level.
When using FACES 2014 data to estimate characteristics of Head Start classrooms (for example, average class size and percentage of classrooms with a teacher who has more than 5 years of teaching experience), you would use the classroom/teacher data file and one of the spring 2015 cross-sectional class weights.

- Teacher survey data alone (T2CLSWT)
- Observation data alone (O2CLSWT)
- Teacher survey data and observation data (TO2CLSWT)

When estimating characteristics of Head Start teachers (for example, average years of teaching experience and highest level of education), you would again use the classroom/teacher data file. Here, you have only one weight to choose from (T2TCHWT). As noted earlier, teachers with more than one sampled classroom have two records in the data file. However, there is not need to remove the duplicate record. T2TCHWT does that for you.
This table shows that records on the classroom/teacher file with a positive value for the class-level weight T2CLSWT and the teacher-level weight T2TCHWT are the same (the file contains a record for each classroom and teachers with two classes have two records). However, the weights sum to different values, because there are more Head Start classrooms than there are Head Start teachers (some teachers have more than one classroom).

<table>
<thead>
<tr>
<th>Weight Name</th>
<th>Unit of Analysis</th>
<th>Description</th>
<th>Records with Positive Weight</th>
<th>Sum of Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2CLSWT</td>
<td>Class</td>
<td>Spring 2015 classroom weight for teacher survey data</td>
<td>614</td>
<td>46,115</td>
</tr>
<tr>
<td>T2TCHWT</td>
<td>Teacher</td>
<td>Spring 2015 teacher weight for teacher survey data</td>
<td>614</td>
<td>41,330</td>
</tr>
</tbody>
</table>
Topic 5.3
Cross-cohort Analysis
FACES is designed to provide national-level information on Head Start and how the program, its services and those it serves change over time. The new Core Plus Study design, which was used for the first time in FACES 2014, places even more emphasis on tracking trends for key indicators.

Because supporting cross-cohort analyses is a critical design consideration, using measures that are comparable across rounds of FACES and limiting changes to the instruments and procedures used to collect data are central features of the FACES design. However, there have been changes to the design of FACES, since the first cohort of children and families was studied in 1997. Appendix A in the User’s Manual summarizes the designs of each of the six rounds of FACES (1997-2014) and is a good place to start to understand the different ways the study has changed, as well as stayed the same.

Researchers who want to use data across multiple rounds of FACES will encounter a number of challenges that are summarized on the slide. First, despite the best efforts to maintain the comparability of measures for key constructs, the measures used do change. This is often done to keep up with advances in the field, and in the case of the child assessments, to use measures that are normed on more contemporary populations of children. Second, small changes can be made to question wording to improve the accuracy of the data collected and to update response options to reflect new programs and practices; however, such changes can make comparisons across rounds of FACES more difficult. Third, the procedures for assessing children who are dual language learners change. For example, the approach and rules for deciding which dual language learners are assessed in English and which are not has changed over the years as have the instruments that are used to assess Spanish-speaking children. Fourth, there can be significant changes to the population of children served by Head Start and to the FACES sample of children who represent this population. As a result, changes in characteristics of children (for example, lower English language skills at program entry) and their families (for example, the percentage of children who live in households where English is not the primary language) may reflect changes in the underlying population (for example, increase in the percentage of children who are Hispanic/Latino).

Researchers will need to carefully review the designs, measures, and procedures used across rounds of FACES and assess the impact they might have on the findings for individual rounds of FACES and for comparisons across rounds.
There are several different approaches that may be used when conducting cross-cohort analyses of FACES data, and the approach used should be driven by the research question. For simple questions such as, has the percentage of children entering Head Start who live with both of their biological parents increased over time or has there been an increase in the percentage of Head Start teachers who have a B.A., generating separate estimates and their standard errors using the data files and sampling weights for each round of FACES might be adequate. This is the approach used in the FACES report, Head Start Children, Families, and Programs: Present and Past Data from FACES that can be downloaded at the ACF/OPRE FACES website: http://www.acf.hhs.gov/opre/resource/head-start-children-families-and-programs-present-and-past-data-from-faces

Another approach that is useful in studying trends and when studying whether subgroup differences are widening, narrowing, or staying the same over time is to combine the data from the different rounds of FACES (for example, FACES 2006 through FACES 2014) into a single data file and use the combined file to examine trends or changes over longer periods. The cases on the combined file should be weighted using comparable weights (named the same) for each round of FACES data and their original strata and PSUs should be employed when estimating standard errors.

For questions about how changes in children’s cognitive and social development are influenced by changes in the composition of the Head Start population, a combined or pooled dataset would be created (for example a dataset that includes all cases with child assessment and parent survey data from FACES 2000 through FACES 2014). A regression framework would be employed to examine the relationships between cognitive and social-emotional outcomes and both child and family characteristics and the year in which the outcomes were measured. Estimates from these regression models would be used to describe changes in mean outcomes over time holding constant the characteristics of the Head Start population. Here again, the data would be weighted using comparable weights and standard errors generated using the appropriate strata and PSUs for each round of data. A variation of this approach has been used to examine trends in the quality of Head Start classrooms using FACES 2006, 2009 and 2014 data: https://www.mathematica-mpr.com/our-publications-and-findings/publications/tracking-quality-in-head-start-classrooms-faces-2006-to-faces-2014-technical-report
Review of Topics 5.1 – 5.3

- To review some of what we have learned about merging data (Topic 5.1), teacher versus class-level analyses (Topic 5.2), and cross-cohort analysis (Topic 5.3), please answer the questions in the Module 5 Review Quiz. Reference slides at the end of this module.

- Once you have answered all the questions, check your answers using the Module 5 Review Quiz Answers. Reference slides at the end of this module.
The FACES 2014 dataset and those of earlier rounds of FACES are complex and can be a bit threatening to users, and especially to users who have more limited experience working with data from large-scale studies with multiple components. This section provides some general guidance for working with the data. The principles outlined in the following slides are intended as an aid and to help ensure that FACES data are used accurately and that the findings from your study are accepted for publication or presentation.
A lot of time was devoted to the preparation of the FACES 2014 User’s Manual. It is your best source for information on the history of FACES and the FACES 2014 design, and for guidance on working with the different FACES 2014 data files. You should carefully read the Manual before beginning any analysis of the data.

It is particularly important that you understand the differences between the two Core studies and the Family Engagement Plus study and which data files to use for different types of research questions. The child file is the only file with child assessment and Core parent survey data. These data as well as any program, staff and observation data on the child file are only for the 60 programs with child-level data collection. The center/program file and the classroom/teacher files include data for all 176 programs (60 programs with child-level data collection and the 116 programs with spring 2015 only data). The Family Engagement Plus study interview data are found on separate data files and the interviews were only conducted in the 60 programs with child-level data collection. However, the supplemental items that were added to the Core parent and Core teacher surveys are located on the child-level file and classroom/teacher file, respectively.

Before you finalize the variables that you will need to answer your research question, carefully review the study questionnaires and codebooks. It is important to know how the questions you are interested in were worded, and whether there were any conditions placed on who was asked a question or set of questions. Respondents may follow different paths through the questionnaires based on their previous participation and some questions are only asked based on the response to a prior question. The codebooks provide information on the unweighted frequency distributions or response ranges for the items and are useful in identifying variables that have limited response ranges and items that might have missing value issues. The codebooks only provide responses for the samples as a whole. If your focus is on a subgroup only (for example, first-time Head Start attendees or minority children) you will want to carefully review the distributions for these smaller samples.

Always check for study-produced constructed or derived variables before you create your own. This will save you time. Using the variables on the data files is one way that you and others can compare findings from different analyses of the FACES data.
You should always examine and evaluate the quality of the variables you are using in your analysis. Pay particular attention to missing data and to outliers. Also, is there a reasonable distribution of responses? And, are the variables related to other variables in the way that you would expect?

Always run your analysis first using unweighted data. This will allow you to more easily identify data with small sample sizes, which can easily be hidden when weights are applied. It will also reveal the extent of variation in the data values.

Once you have completed your preliminary analysis using the raw (unweighted) data, all analyses should use the sampling weight that is appropriate for your research question and the data that are being used. The weights adjust for any differential probabilities of being included in the study samples and for unit nonresponse, and thus help to reduce the changes that your results will be biased.
Use the design-based method (Taylor Series) when estimating variances and standard errors whenever possible. This helps to guard against underestimating standard errors and introducing errors into your findings based on inflated significance test results. If you do not have access to software that supports the Taylor Series method, use the approximation method to adjust for the design effects that was described in Module 4.

The FACES sample is large and it can sometimes be easy to find significant results when comparing certain groups (for example, children attending Head Start for one versus two years) because of this. When reporting findings, don’t just focus on whether group differences or coefficients are statistically significant. Include information about the magnitude of the difference (effect size). Is the difference meaningful or is it merely significant as a function of a large sample size? Researchers often convert the observed difference into an effect size and use this to interpret the size of the difference. For simple t-tests, the effect size is the observed difference between two means relative to a measure of variability, such as the standard deviation. In correlation and regression analyses, $r$ and $B$ are the effect size, respectively. A good introduction to effect size and how it is measured when using different statistical procedures can be found in Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press.

Finally, always consider the plausibility of your results. Do the results make sense given what you know about Head Start and what you know from your knowledge of the research literature? Findings that are at odds with what you would expect should be closely scrutinized. Assume that there is a mistake in your analysis until you have examined and eliminated all of the possible reasons for the unlikely finding.
Each of the topics covered in this module is introduced in the Getting Started preface to the FACES 2014 User’s Manual with more discussion in Chapter VI:

• Merging data from different files: See Chapter VI, Section A.6, pg. 177
• Teacher- and classroom-level analysis: See Chapter VI, Section C, pg. 189
• Cross-Cohort Analysis: See Chapter VI, Section D, pg. 192

Appendix A contains a summary of key features of the FACES design and how it has changed and stayed the same over the years. More details on the features of each round of FACES can be found in the FACES 1997 – 2009 user’s manuals. Many of the topics covered in the module are discussed in earlier FACES manuals.
MODULE 5
REVIEW QUIZ & ANSWERS
Module 5 Review Quiz

1. All child-level analyses using FACES 2014 can be completed using a single file – the child file. This file contains all of the data that were collected from all of the data sources.
   a. True
   b. False

2. Cases with missing value codes should always be eliminated from your analysis.
   a. True
   b. False

3. There are several unique features of the classroom/teacher file that should be considered when estimating the characteristics of Head Start teachers. Which of the following are features of that file? (Check all that apply)
   a. The number of classroom records and teacher records on the file are the same
   b. Each teacher has a single record
   c. There are records on the file for classrooms and teachers who were not originally part of the FACES 2014 classroom and teacher samples
   d. All of the above
Module 5 Review Quiz, Cont’d

4. Which of the following pose challenges to researchers using data from across rounds of FACES to study changes and trends in Head Start? (Check all that apply)
   a. Measures change sometimes to keep up with advances in the field and to capture new program priorities
   b. Small changes have been made to question wording
   c. There have been changes to the procedures and measures that are used for assessing children who are dual language learners
   d. There can be significant changes to the population the program serves

5. One approach to cross-cohort analysis is to combine the data from the rounds of FACES (for example, FACES 2003 – 2014) into a single file and use that file to study changes and trends. Researchers using this approach would need to develop a “mega weight” that would represent the population of children who attended Head Start over the years of interest (2003 – 2014) and use this weight in all of their analyses.
   a. True
   b. False
Module 5 Review Quiz Answers

1 ANSWER: False. While the child file does contain data from the different data sources, the only item-level data found on this file come from the parent survey and the Teacher Child Report (and this is only a subset of the teacher ratings items that were not redacted). You will need to merge item-level data from the other Core data files if your research question requires that you use item-level data from these instruments.

2 ANSWER: False. Whether or not to exclude these cases depends on your research question and the statistic you are generating. There are instances when cases are set to missing because of a skip pattern in the questionnaire, which is indicated by a missing value code of -1 (SPSS) or N (SAS) in the data file. You may need to recode these to a valid value depending on the reason for the item being skipped. For example, parents may only be asked to name the country where they were born if they first indicated that they were born outside of the U.S. in a prior question. Estimating the percentage of all parents who were born in Mexico without recoding the -1 (or N) missing value codes would result in an overestimate.

3 ANSWER: The correct answers are a and c. There are a total of 691 records on the file (one for each classroom) and teacher background data are merged onto the classroom record (a). The classroom/teacher file includes a record for each classroom that was sampled and a record for 24 classrooms and their teachers that children moved into. They were not part of the classroom sample.
Module 5 Review Quiz Answers, Cont’d

4  ANSWER: All of the answer options are correct (a-d). Measures change or replace earlier versions (for example, the version of the PPVT used in FACES has changed several times over the years). Question wording can be tweaked and response options added to improve measurement of the underlying construct. The cut score for deciding who is assessed in English and who is not has changed and a new measure of Spanish receptive vocabulary (ROWPVT-SBE) has replaced an old standby (TVIP). The racial/ethnic composition of the Head Start population has changed over the years with larger percentages of Hispanic/Latino children and families participating. This demographic change can drive other changes in the characteristics of children and their families, and in the ways in which programs offer services.

5  ANSWER: False. Researchers using this combined file approach would weight the data for each case using a comparable weight from each of the FACES rounds. For example, researchers interested in studying trends in the school readiness skills of children when they first enter Head Start would use a weight from each of the fall waves of data collection that is designed to be used when analyzing child assessment data (and perhaps in combination with the parent survey data).