Toddlers in Early Head Start: A Portrait of 3-Year-Olds, Their Families, and the Programs Serving Them

Volume I: Age 3 Report

April 2015

Baby FACES 2009

OPRE Report 2015-28

This page is left blank for double-sided printing.

Toddlers in Early Head Start: A Portrait of 3-Year-Olds, Their Families, and the Programs Serving Them

Volume I: Age 3 Report
OPRE Report 2015-28

April 2015

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

Submitted by:
Cheri A. Vogel
Pia Caronongan
Yange Xue
Jaime Thomas
Eileen Bandel
Nikki Aikens
Kimberly Boller
Lauren Murphy
Mathematica Policy Research

Project Director:
Cheri A. Vogel
Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393

Contract Number: HHSP23320072914YC
Mathematica Reference Number: 06432.150

Suggested citation:
Vogel, C.A., P. Caronongan, Y. Xue, J. Thomas, E. Bandel, N. Aikens, K. Boller, and L. Murphy (2015). Toddlers in Early Head Start: A Portrait of 3-Year-Olds, Their Families, and the Programs Serving Them. OPRE Report 2015-28. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

MATHEMATICA Policy Research

baby FACES

OPRE

This page is left blank for double-sided printing.

TECHNICAL WORK GROUP MEMBERS

We would like to thank the following members of the Baby FACES Technical Work Group. The views expressed in this publication do not necessarily reflect the views of these members.

Catherine Ayoub
Harvard University

Lisa Berlin
University of Maryland

Jeanne Brooks-Gunn
Teachers, College
Columbia University

Margaret Burchinal
University of North Carolina
Frank Porter Graham Institute

Judith Carta
University of Kansas

Michael Foster*
University of Alabama at Birmingham

Carol Hammer
Temple University

Brenda Jones Harden
University of Maryland

Judith Jerald
Save the Children

Ellen Eliason Kisker
Twin Peaks Partners, LLC

Helen Raikes
University of Nebraska

Lori Roggman
Utah State

Tammy Mann
Campagna Center

*We are very sad to note that Michael Foster died before publication of this report.

This page is left blank for double-sided printing.

CONTENTS

TABLES

This page is left blank for double-sided printing.

FIGURES

This page is left blank for double-sided printing.

EXHIBITS

This page is left blank for double-sided printing.

OVERVIEW

The Early Head Start Family and Child Experiences Survey (Baby FACES) is a descriptive study of Early Head Start programs designed to inform policy and practice at both national and local levels. In 2007, the Office of Planning, Research & Evaluation (OPRE) in the Administration for Children and Families (ACF), U.S. Department of Health and Human Services, contracted with Mathematica Policy Research and its partners to implement this longitudinal study in 89 Early Head Start programs.

Baby FACES followed two cohorts of children through their time in Early Head Start, starting in 2009, the first wave of data collection. The Newborn Cohort includes 194 pregnant mothers and newborn children. The 1-year-old Cohort includes 782 children who were approximately age 1 (ranging from 10 to 15 months). This third and final report focuses on the last waves of data collection and children who were 3 years old in 2011 and 2012. We present findings on the three aims of Baby FACES:

1. **Describing Early Head Start and program services and staff**

- On average, observed quality of both classrooms and home visits is in the mid-range:
  - Home visit quality is positively associated with the amount of time spent on parent-child activities, the alignment of the visit with the visitor's plan, and the presence of another adult during the visit. It is negatively associated with time spent on staff-parent relationship building.
  - Classroom quality is positively associated with teacher job satisfaction and the quality of teacher-parent relationships. It is negatively associated with teacher depressive symptoms (Emotional and Behavioral Support domain only).

2. **Describing change over time in the population served by the program**

- During enrollment, parents and children show improvements in several domains:
  - Parents report better mental health and lower stress over time.
  - Children's English vocabulary skills increase in a nonlinear fashion with a rapid increase at age 1 that slows down between ages 2 and 3.
  - Children's social-emotional skills improve over time.

3. **Relating program services to child and family outcomes**

- Looking at associations between service take-up and quality and child outcomes at age 3 reveals a few positive relations (although many more nonsignificant ones):
  - Children in families rated as highly involved have better behavioral outcomes.
  - The number of center days children attend is positively associated with auditory comprehension.
  - Among Spanish-speaking children, home visit quality is positively related to a Spanish measure of auditory comprehension.
  - Engaged Support for Learning, an aspect of classroom quality, is positively related to children's receptive vocabulary.

Next Steps/Looking Ahead

A series of short reports and program-friendly briefs will address other topics of interest.

This page is left blank for double-sided printing.

EXECUTIVE SUMMARY

The Early Head Start Family and Child Experiences Survey (Baby FACES) is a descriptive study of Early Head Start programs designed to inform policy and practice at both national and local levels. In 2007, the Office of Planning, Research & Evaluation (OPRE) in the Administration for Children and Families (ACF), U.S. Department of Health and Human Services, contracted with Mathematica Policy Research and its partners to implement this longitudinal study in 89 Early Head Start programs around the country. Baby FACES followed two cohorts of children, newborns and 1-year-olds, through their time in Early Head Start. The Newborn Cohort includes pregnant mothers and newborn children (194 are in this group) and the 1-year-old Cohort includes children who were approximately age 1 (782 were aged 10 to 15 months) at study enrollment in 2009. Data collection started in the spring of 2009 and ended for the 1-year-old Cohort in spring 2011 and for the Newborn Cohort in spring 2012, when both cohorts were 3 years of age. This is the third and final report describing the experiences of families and children in Early Head Start. The first report provides in-depth information about the sample design, the measures used, and the baseline findings (Vogel et al. 2011) and the second report describes findings from the second wave of data collection focused primarily on children who were 2 years old in 2010 (1-year-old Cohort only) (Vogel et al. 2015). This report describes the experiences of children in both cohorts through age 3 and focuses on understanding program participation and predictors of participation, service quality and predictors of quality, and associations between receiving services at different levels of intensity and quality and child and family outcomes.

Research questions for Baby FACES address three primary aims: (1) describing Early Head Start and program services and staff, (2) describing the population served by the program, and (3) associating program services with child and family outcomes. Specific questions addressed by this report include:

- What is Early Head Start? What are the program models employed, staff qualifications, and other important program features and characteristics?
- What specific services are delivered to families and what is their quality?
- What are the characteristics of the families Early Head Start serves in terms of their demographic, household, and family characteristics; their needs; and their risk factors?
- How are Early Head Start children and families faring over time?
- What are the predictors of home visit and classroom quality?
- What are the predictors of program participation?
- How do family characteristics, program experiences, and quality relate to outcomes?

Box 1 includes brief information on the data sources and measures used at age 3 and Table II.1 provides additional detail.

Box 1. Overview of Baby FACES Data Sources at Age 3

Parent Interview. This telephone interview asked the person primarily responsible for the care of the study child about demographic characteristics, their service needs and use, and their well-being and that of the child. It also asked about the child's exposure to environmental health risks and environmental and routine supports for the child's growth and development. Parents were also asked to rate their child's development and behavior.

Direct Child Assessment and Home Observation. The assessments include administration of the Preschool Language Scale-4 Auditory Comprehension subscale (PLS-4; Zimmerman et al. 2002), the Peabody Picture Vocabulary Scale–Fourth Edition (PPVT-4; Dunn and Dunn 2007), and measurement of height and weight. While in the home, the field assessor also observed the child's ability to focus on the tasks, the interactions between the child and parents, and the quality of the home environment using the Bayley Behavior Rating Scale (BRS) (Bayley 2006), the Home Observation for Measurement of the Environment (HOME) (Caldwell and Bradley 2003), and scales drawn from a study of neighborhoods in Chicago (Ross et al. 2008).

Parent Self-Administered Questionnaire. Parents were asked to rate their child's development and behavior using the Ages & Stages Questionnaires, Third Edition [ASQ-3] (Squires et al. 2009), MacArthur-Bates Communicative Development Inventories [CDI] (Fenson et al. 2000), the Brief Infant Toddler Social Emotional Assessment [BITSEA] (Briggs-Gowan and Carter 2006), and the Behavior Problems Index (Zill and Peterson 1986). They also rated the quality of their relationship with the child's home visitor or teacher.

Parent-Child and Assessor-Child Interaction. Children participated in two semi-structured interaction activities that involve playing with two different sets of toys. Parents interacted with children using the Two-Bag task protocol, an adaptation of the parent-child interaction task used in the Early Head Start Research and Evaluation Project (EHSREP) and the Early Childhood Longitudinal Study-Birth Cohort (ECLS-B). The videos were coded using two coding schemes: Parent-Child Interaction Rating Scales for the Two-Bag Assessment (Mathematica Policy Research 2010) and an adaptation of the Parenting Interactions with Children: Checklist of Observations Linked to Outcomes (PICCOLO; Roggman et al. 2009). The field assessors also interacted with children following the Early Communication Indicator (ECI) protocol (Greenwood et al. 2006), that was later scored from videos.

Staff-Child Report. Home visitors and teachers of study children completed child-specific ratings of the participation of families in the program, the behavior of children using the Behavior Problems Index (BPI; Zill and Peterson 1986), and the quality of their relationship with the parents of study children.

Home Visitor/Teacher Interview. We interviewed either the child's home visitor or the child's teacher to determine her demographic characteristics, tenure working for the program, and well-being, as well as training and education experiences provided by the program and the work environment.

Classroom Quality Observation. Observers rated classrooms with the CLASS-T (Pianta et al. 2010) which measures the quality of teacher-child interactions in center-based settings and includes two subscales: Engaged Support for Learning and Emotional and Behavioral Support.

Home Visit Quality Observation. Field assessors observed the home visitors who provided services to children in the study sample using the Home Visit Rating Scales-Adapted (HOVRS-A) (Roggman et al. 2009) and a form that assessed the content and characteristics of the visit.

Family Services Tracking (FST). Early Head Start home visitors and teachers of study children completed a weekly service tracking form that detailed the number of service experiences (home visits or days in care) study children were offered and the number received.

What Are the Program Models Employed, Staff Qualifications, and Other Important Program Features and Characteristics?

Prior reports provide a more detailed answer to this question (see Vogel et al. 2011 and Vogel et al. 2015). At age 3, information was collected on staff qualifications and experience.

Staff Have Extensive Early Head Start Experience

Services are provided by diverse and experienced teachers and home visitors, usually in children's home language. Baby FACES children are served by teachers and home visitors who usually provide services in the child's home language—78 percent of children from Spanish speaking homes have a home visitor who speaks Spanish, and 82 percent of such children in center-based services have a teacher or another adult Spanish speaker in the classroom. Children's staff have on average 6 to 7 years of experience in Early Head Start and some have a college degree (37 percent of children have teachers with a B.A. or higher and 59 percent of children had home visitors with this level of education). Further, staff reported positive feelings about their jobs and low levels of depressive symptoms and experienced low turnover rates.

What Specific Services are Delivered to Families and What is Their Quality?

Programs Deliver Services of Mid-range Quality

Home visit quality is in the mid-range. Using the 5-point HOVRS-A observational measure, home visit quality scored in the mid-range both overall (mean total score 3.4), and on two subscales focused on engaging the family in the visit (Visitor Effectiveness; mean 3.8), and in their responsiveness and skill in facilitating parent-child interaction (Visitor Strategies; mean 3.2). Most of the time during visits was spent on a variety of child-focused activities (57 percent of the visit on average).

Classroom quality is moderate. Observed quality scores on the 7-point CLASS-T are highest in the area of Emotional and Behavioral Support (5.3), and lowest in the area of Engaged Support for Learning (3.3).

How are Early Head Start Children Faring over Time?

Children are Faring Well and Showing Improvement over Time

The Baby FACES study team followed children and families throughout their enrollment in Early Head Start, collecting data each year through age 3 by observation, parent and staff report, and direct assessments with standardized instruments. Although there was some sample attrition over the course of the study (approximately one-third of all families exited the program before children were 3—See Box 2), we did not see large differences in the characteristics of those who left the program early compared to those who stayed. (See Appendix D.) Overall we found that children are making good developmental progress by most measures.

Parents rate children highly on physical health, general development, and language. Parents report that their children are in very good or excellent health as well as have access to regular health care and high rates of insurance coverage. However, similar to national rates, measurements of height and weight show that overweight and obesity are a concern, with about one-third of the children falling into one of these groups by age 3. Parents rate children's speaking ability, vocabulary and comprehension of their home language as strong.

Box 2. Baby FACES Sample, Response Rates, and Analytic Approaches

In spring 2009 there were 976 children who were eligible and consented to be in the study. By the time the children were three years old, 253 families out of 782 originally enrolled in the 1-year-old Cohort left Early Head Start and 109 families out of 194 originally enrolled in the Newborn Cohort did so. Children who left the Early Head Start program from which they were sampled were considered no longer eligible at follow-up, and this was by far the main driver of sample attrition over time.

Eligible and Consented Sample Sizes at Baseline and Follow-Up by Year

Cohort	2009	2010	2011	2012
Newborn	194	140	100	85
1-Year-Old	782	602	469	n.a.
Combined	976	742	569	85

Response Rates Are High Across Study Waves

Despite difficulties in maintaining high response rates in a longitudinal study, the study achieved high rates of completion for each of the data collection instruments over time. The Family Services Tracking (FST) data analyzed for this report covers a 104-week period from the spring of each child's age 1 year to the spring of his/her age 3 year. On average, from age 1 to age 2, reports were submitted for 73 percent of eligible weeks. From age 2 to 3, reports were submitted for 69 percent of eligible weeks.

Baby FACES Response Rates 2009–2012

Instrument	2009 Number Completed (Percentage)	2010 Number Completed (Percentage)	2011 Number Completed (Percentage)	2012 Number Completed (Percentage)
Staff-Child Report	933 (95.5)	703 (95.6)	538 (96.2)	82 (97.6)
Parent Interview (CATI)	894 (91.7)	583 (79.3)	445 (79.6)	61(72.6)
Parent Self-Administered Questionnaire (SAQ)	n.a.	537 (89.5)	481 (86.0)	70 (83.3)
Child Assessment	n.a.	547 (91.2)	503 (90.0)	76 (90.5)
Caregiver Interview	229 (93.1)	267 (98.9)	232 (98.7)	44 (100)
Home Visitor Interview	323 (96.7)	225 (97.0)	174 (99.4)	29 (100)
ITERS-R	223 (94.9)	53 (98.1)	n.a.	n.a.
CLASS-T[b]	n.a.	220 (98.7)	231 (99.1)	42 (95.5)
HOVRS-A	242 (89.3)	193 (83.2)	139 (84.2)	20 (87.0)
Program Director Interview	89 (100)	89 (100)	89 (100)	n.a.
Program Director SAQ	86 (96.6)	83 (93.3)	n.a.	n.a.
Exit Interview[a]	62 (54.9)	76 (38.3)	98 (57.9)	337 (72.8)

Source: Baby FACES Sample Management System (SMS).

Note: Percentages are of those still enrolled in Early Head Start and therefore still eligible for the study at each wave. Exit interviews are of those who were reported to have left the program at each time period.

[a]Administered to parents of children who left the program by each round of data collection. Round 1 was administered between October and December 2009. Round 2 was administered between April and June 2010; Round 3 was administered February to June 2011. The total number of cases released for Round 2 includes the 51 incomplete responses from Round 1. Overall, between Rounds 1 and 2, we released a total of 258 unique cases. The combined response rate of Rounds 1 and 2 is 54 percent. The final round was administered to parents of 1-year-old Cohort children who remained enrolled through age 3, when the children were 3 1/2 years old (August through October 2011).

[b]Administered in classrooms of 2- and 3-year-olds. The ITERS-R was used to rate classrooms of 1-year-olds.

ITERS-R = Infant Toddler Environment Rating Scale-Revised; HOVRS-A = Home Visitor Rating Scale-Adapted; CLASS-T=Classroom Assessment Scoring System-Toddler version; n.a. = not applicable.

Analytic Approaches

In this final report we use data on children's entire experience in Early Head Start to examine longitudinal development and relationships among program and family characteristics, Early Head Start experiences (participation in and quality of services), and outcomes. We use simple descriptive statistics such as

means/percentages and standard errors (weighted as appropriate) to depict 3-year-old children and their families in our sample. For analysis of growth over time, predictors of home visit and classroom quality, and predictors of Early Head Start participation we created Hierarchical Linear Models (HLM). To analyze relationships between participation and outcomes we first assess the bivariate relationship between each participation and outcome measure. For significant associations, we proceeded with multivariate analyses that include a comprehensive set of child, family, and program characteristics as covariates. Detailed information of the analytics approached are provided in Chapter II, Chapter IX, and Appendix D.

Children are continuing to develop their language skills at age 3 but are not quite at national norms. Direct child assessment suggests that Early Head Start children's auditory comprehension as measured by the PLS-4 is approaching national norms, while receptive vocabulary as measured by the PPVT-4 and expressive language skills as measured by the ECI have some catching up to do with their same-aged peers.

When looking over the course of children's participation in the study, we see nonlinear growth in children's English vocabulary with a rapid increase at age 1 that slows down between ages 2 and 3, following a similar pattern as a normative sample on the CDI. English vocabulary growth trends of dual language learners (DLLs) are similar to children from English speaking homes, although these children display a smaller English vocabulary, on average. When just focusing on Spanish speaking children assessed using the Spanish CDI, we see linear Spanish vocabulary growth, indicating a consistent rate of growth over time.

Although multiple data sources present a mixed picture of children's social-emotional development at age 3, children's social-emotional skills improve over time. Parents report significantly more social-emotional problems for children than do Early Head Start staff based on the BITSEA; however, parents report slightly but significantly fewer behavior problems for children than do Early Head Start staff based on the BPI. Assessor observations of children's task engagement and emotional regulation are approaching national norms. Additionally, the majority of children display some positive behaviors in video-recorded play interactions with their parents on the Parent Child Interaction Rating Scales and display few negative behaviors. Estimates are comparable to those reported in other large-scale studies with children of similar ages.

When looking at children's growth from age 1 to age 3, parent reports of social-emotional competence on the BITSEA (empathy, prosocial behavior, and compliance) increased in a linear fashion over time.

Parents provide supportive and cognitively stimulating environments, although often families reside in neighborhoods with poor conditions. On average, the home environments that parents provide are rated as emotionally supportive, cognitively stimulating, and well organized as measured by the HOME. Parents provide high levels of emotional support and rarely show harsh or punitive parenting behaviors. Additionally, on average, parents provide environments with cognitively stimulating books, toys, and other materials and assessors rate the interior of the homes as generally clean and well organized. In contrast to the interior of the homes, observers rate the neighborhoods in which families live as being in poor condition, indicating environments that in general have rundown housing, are strewn with litter and other trash, and have an unsafe atmosphere.

Parents report better mental health and lower stress over their enrollment in Early Head Start. Parent depressive symptoms on the CESD-SF and levels of parenting stress on the PSI-SF

improve over enrollment in Early Head Start. Depressive symptoms decrease in a nonlinear manner, with the largest decrease between birth and age 2. Parenting stress declines in linear fashion.

What Are the Predictors of Home Visit and Classroom Quality?

Family, Staff, and Program Characteristics Associated with Quality Differ for Home and Classroom Services

We constructed multi-level hierarchical linear models (HLM) to examine whether the home visit or classroom quality provided by a particular home visitor or teacher changes over time (between 2009 and 2012 for home visit quality and 2010 and 2012 for classroom quality). We also examine whether particular staff or program characteristics are associated with quality (see Chapter II for additional details on the analytic approach).

Most aspects of quality are stable over time. For home visiting, the quality provided by home visitors is stable over time. For classrooms, we observe different patterns of quality provided by teachers across the two classroom quality domains: Emotional and Behavioral Support and Engaged Support for Learning. Specifically, we observe stability in Emotional and Behavioral Support and a decline in Engaged Support for Learning over time.

Home visit content and characteristics are associated with home visit quality over time, but few staff or program characteristics are associated with quality. Specifically, the amount of time spent on parent-child activities, the alignment of the visit with the visitor's plan, and the presence of another adult during the visit are all positively associated with quality as indicated by both subscales of the HOVRS-A, Visitor Strategies and Visitor Effectiveness. Conversely, time spent on staff-parent relationship building is associated negatively with the home visit quality subscales. There are some differences by subscale. For Visitor Strategies there is a positive relationship with job satisfaction, and for Visitor Effectiveness there is a negative association with family focused activities. No other staff or program characteristics related to home visit quality, such as program approach and implementation or staff race/ethnicity, credentials, and language.

Only one child and family characteristic is related to home visit quality over time. Specifically, visits conducted with younger children receive lower quality ratings on both Visitor Strategies and Visitor Effectiveness compared to visits conducted with older children.

Staff characteristics and parent-teacher relationships are associated with classroom quality over time. Some staff characteristics are associated with classroom quality, while program-level characteristics such as program approach, population served, and implementation, are not. Specifically, teacher educational level is associated with classroom quality (Engaged Support for Learning only). Teacher job satisfaction and the quality of teacher-parent relationships are positively associated with classroom quality (Emotional and Behavioral Support and Engaged Support for Learning). Teacher depressive symptoms are negatively associated and percentage of DLLS in the classroom is positively associated with Emotional and Behavioral Support.

What Are the Predictors of Program Participation?

The dimensions of program participation we examined are length of enrollment, staff ratings of family involvement, and service take-up rates (percentage of services offered by programs that were received by families).

Families Enroll for over 2 Years On Average, and Are Highly Engaged While Children Are Young.

Children and families enroll for over 2 years on average, varying by timing of entry. Among all children in the sample, the average length of enrollment is 28 months. However, the average varies by the timing of enrollment. Among families who enrolled during pregnancy, the average length of enrollment is 33 months. In contrast, the average length of enrollment among children whose families enrolled after birth (and by 15 months of age) is 25 months. For families who left before age 3 the average length of enrollment is 17 months.

Staff rated families as demonstrating high involvement at ages 1 and 2. At ages 1 and 2, the majority of families are rated by staff as highly involved. However, by age 3, the percentage of families demonstrating high involvement (among those still enrolled) declined and more families are rated as having inconsistent involvement.

Programs offer home visits and center days at high frequencies. The study asked program staff to track service provision on a weekly basis when children were between 1 and 3 years of age and enrolled in the program. On average, families in the home-based option are offered approximately 1 home visit per week when children are between ages 1 to 3. The average total number of home visits offered each year ranges from 44 to 48 (the OHS recommends programs offer 48 visits per year). The average number of center days offered to families in the center-based option is approximately 4 per week. The average number of center days offered per year is 224 between ages 1 and 2 and 192 days between ages 2 and 3 (the OHS recommends programs offer 240 center days per year).

Service take-up by families is high, on average, but varies by service delivery type. Take-up is defined by the number of home visits and center days offered by programs divided by those received by families over a two-year period. As part of the study, service provision and family receipt were tracked on a weekly basis by program staff when children were between 2 and 3 years of age and enrolled in the program. Rates of take-up differ by the service delivery type (home visits versus center-based care). Families in the home-based option for a full year complete 77 percent of the home visits offered between ages 1 and 2, on average, and 75 percent between age 2 and 3, on average. Children who are in the center-based option for a full year attend 85 percent of center days offered between ages 1 and 2, on average, and 86 percent between ages 2 and 3, on average. Children who left the program before age 3 have lower home visit and center day take-up rates while they are enrolled.

Program Participation Varies by Child and Family Characteristics

We constructed multi-level HLM models to examine participation in Early Head Start as defined by three indicators of participation: length of enrollment, staff ratings of family involvement, and service take-up rate.

Participation does not vary substantially between programs, regardless of the measure of participation. Programs included in Baby FACES are roughly equivalent in levels of family participation, with most of the differences explained by family-level factors. We observe a great deal of within-program variability in family participation. For example, 78 percent of the variability in length of enrollment is due to differences between families. Similarly, half of the variability in involvement ratings within a given program is due to differences between families. Families also differ considerably in terms of their service take-up rates within a given program (more than half of the variability in home visit and center take-up rates is due to between-family differences).

Length of enrollment does not vary by many child, family, or program characteristics. Children who are DLLs stay in Early Head Start longer than children from homes in which English is the

only language spoken. There are no significant associations between length of enrollment and other child and family characteristics, program approach, population served, or program implementation.

Family involvement is associated with some child and family characteristics, but is not associated with program characteristics. Families are more likely to be rated by staff as highly involved when their children are younger (age 1) compared to when their children are older (age 2 and age 3). Families who left the program early are less likely to be rated as highly involved while enrolled compared to families who stay through age 3. Families with medium or high levels of demographic risks are less likely to be rated as highly involved compared to families with lower risks. Family involvement is not related to program approach, population served, or program implementation.

Home visit take-up rates vary by child age while center take-up rates do not. Specifically, we observed that home visit take-up rates between ages 1 and 2 are higher compared to take-up rates between ages 2 and 3. There are no analogous differences by age for center take-up rates.

Child and family language and race/ethnicity are associated with home visit take-up rates. Take-up rates are higher for both home visits and center days for those who stay in the program until age 3 relative to early leavers (while they were enrolled). Similarly, DLLs have higher home visit take-up rates than families who speak English only. However, when looking at home visit take-up by race/ethnicity, Hispanic[1] families have lower home visit take-up rates relative to white families.

Staff race/ethnicity and education but not program characteristics are associated with home visit and center take-up rates. Specifically, families whose home visitors are African American have higher home visit take-up rates compared to those with white home visitors. Children whose teachers have at least a bachelor's degree have higher center take-up rates than children whose teachers have lower levels of education. Program characteristics such as program approach, population served, and implementation are not related to service take-up.

How Do Family Characteristics, Program Experiences, and Quality Relate to Outcomes?

Finally, we examined how factors such as family participation in the program, the quality of the services received over time, and child, family, staff, and program characteristics might be associated with key child and family outcomes.

Some Forms of Program Participation (Family Involvement, Length of Enrollment, and Take-up) Are Associated with Child Outcomes While Others Are Not

- ***Children in highly involved families have better behavioral outcomes at age 3.*** Controlling for a large set of child, family, and program characteristics, children of consistently highly involved (over two years) families exhibit significantly fewer negative and significantly more positive behaviors as rated by their teacher or home visitor using the BITSEA compared to children of less involved families. Children in these families also have better emotional regulation capabilities as rated by observers on the BRS.
- ***Center day take-up is positively associated with age 3 language abilities.*** Total center days attended are positively associated with the children's auditory comprehension as measured by the English PLS-4.

[1] Throughout this report, Hispanic refers to those with a Hispanic or Latino racial/ethnic background.

- ***Enrolling during pregnancy, length of enrollment, and home visit take-up are not associated with age 3 child and family outcomes.***

Associations Between Service Quality and Child Outcomes Are Mixed

- ***Average home visit quality is not associated with age 3 outcomes, but quality thresholds may matter for Spanish language development.*** Spanish speaking children with higher-quality home visits (i.e., those with an average score of 3 or more on the HOVRS-A Visitor Strategies subscale) have significantly higher Spanish PLS-4 scores at age 3.
- ***Center quality is positively associated with age 3 language abilities.*** We find a positive, statistically significant association between the CLASS-T Engaged Support for Learning subscale and receptive vocabulary as measured by the PPVT-4 at age 3.

Next Steps/Looking Ahead

This report is the final in a series of three that describes the children and families who participated in the Baby FACES project. Baby FACES provides a comprehensive longitudinal descriptive look at the Early Head Start program, staff, services, and families and examines the associations among these different factors. In addition to the set of three longer reports, a series of short reports and program-friendly briefs will address topics of interest in more depth and with an eye towards informing local program planning and use of data.

This page has been left blank for double-sided copying.

I. EARLY HEAD START RESEARCH AND THE BABY FACES STUDY

The Early Head Start Family and Child Experiences Survey (Baby FACES) is an ongoing study of Early Head Start programs designed to inform policy and practice at both national and local levels. In 2007, the Office of Planning, Research & Evaluation (OPRE) in the Administration for Children and Families (ACF), U.S. Department of Health and Human Services (USDHHS), contracted with Mathematica Policy Research and its partners to implement this six-year longitudinal study in 89 Early Head Start programs around the country. We enrolled two cohorts of children into the study in spring 2009: (1) a Newborn (less than two months old) Cohort and (2) a 1-year-old Cohort, comprising children who were approximately age 1 at our first wave of data collection. We followed children and families each spring until they turned 3 or left the program. See the Baby FACES baseline report (Vogel et al. 2011) for a full description of the sample at enrollment in the study and the follow-up report describing children when they were 2 years old (Vogel et al. 2015). This final report describes children at age 3. Chapter I describes the context in which Early Head Start is operating, gives an overview of the study, provides highlights of findings at age 2, and describes new features of the data collection effort in this wave. The chapter ends with a road map to the report.

The Early Head Start Program

Early Head Start is a two-generation program that began in 1995 as a federal initiative for low-income pregnant women and families with infants and toddlers 3 years old or younger. Early Head Start programs provide a wide range of services, including child development services, child care, parenting education, case management, health care and referrals, and family support. In addition to delivering many services directly, programs also form partnerships with other community service providers to meet families' needs.

To ensure the quality of their offerings, Early Head Start programs adhere to two key institutional benchmarks: the Head Start Program Performance Standards (hereafter performance standards) and the Framework for Programs Serving Infants and Toddlers and Their Families. The performance standards are the rules and regulations that explicitly identify what programs must do to ensure high-quality services (for example, they specify child-to-adult ratios in child care centers, educational requirements for staff, and the types of services that must be offered) (ACF 1996). The framework is a conceptual model that describes the mechanisms by which high-quality programs are thought to affect children's outcomes. The framework is structured as a pyramid that rests on a foundation of four cornerstones (community, staff, family, and child development) that the Advisory Committee on Services for Infants and Toddlers deemed essential for quality Early Head Start programs (ACF 1994). On this foundation, the framework builds four layers; management systems form the base that supports program services. These services bring about positive family and child outcomes and contribute to the goal of children's well-being and competence (Figure I.1).

The performance standards define five service delivery options that programs can use based on the unique needs of families:

1. Home-based—families receive weekly home visits and at least two group socializations per month
2. Center-based—families receive center-based child care plus other activities
3. Combination—families receive both home visits and center-based experiences
4. Locally designed—requiring official approval from the Office of Head Start (OHS)

5. Family child care—families are served through family child care homes

A program can choose to deliver one option to all families or different combinations to different families, based on its determination of the best mix of services for meeting families' needs.

Early Head Start has expanded over time. From the initial 68 Early Head Start grantees funded in 1995, the program had by 2009 grown to more than 700 programs serving more than 60,000 children and families (Early Head Start National Resource Center 2010). In a large-scale policy shift, the federal government allocated $1.1 billion for the expansion of Early Head Start through the American Reinvestment and Recovery Act of 2009 (ARRA), adding nearly 50,000 slots in fiscal year 2009–2010 (ACF 2010) and reversing prior polices that provided a flat funding stream without increases for inflation or the addition of new enrollment slots. As a result of the ARRA expansion, nearly 1,000 programs now serve about 100,000 children. The Baby FACES sample includes 54 ARRA expansion grantees out of the 89 sampled programs. In April 2011, $340 million in funding was allocated to continue operations for the ARRA expansion programs, to be received in September 2011 for a six month continuation of those services. In December 2011, the Consolidated Appropriations Act of 2012 was signed increasing funding by an additional $69 million for the continuation of the expansion programs, a cost-of-living adjustment, and additional technical assistance and training. Expansion programs that received ARRA funding will have the funds added to their base funding (http://eclkc.ohs.acf.hhs.gov/hslc/standards/PIs/2012/resour_pri_001_012612.html).

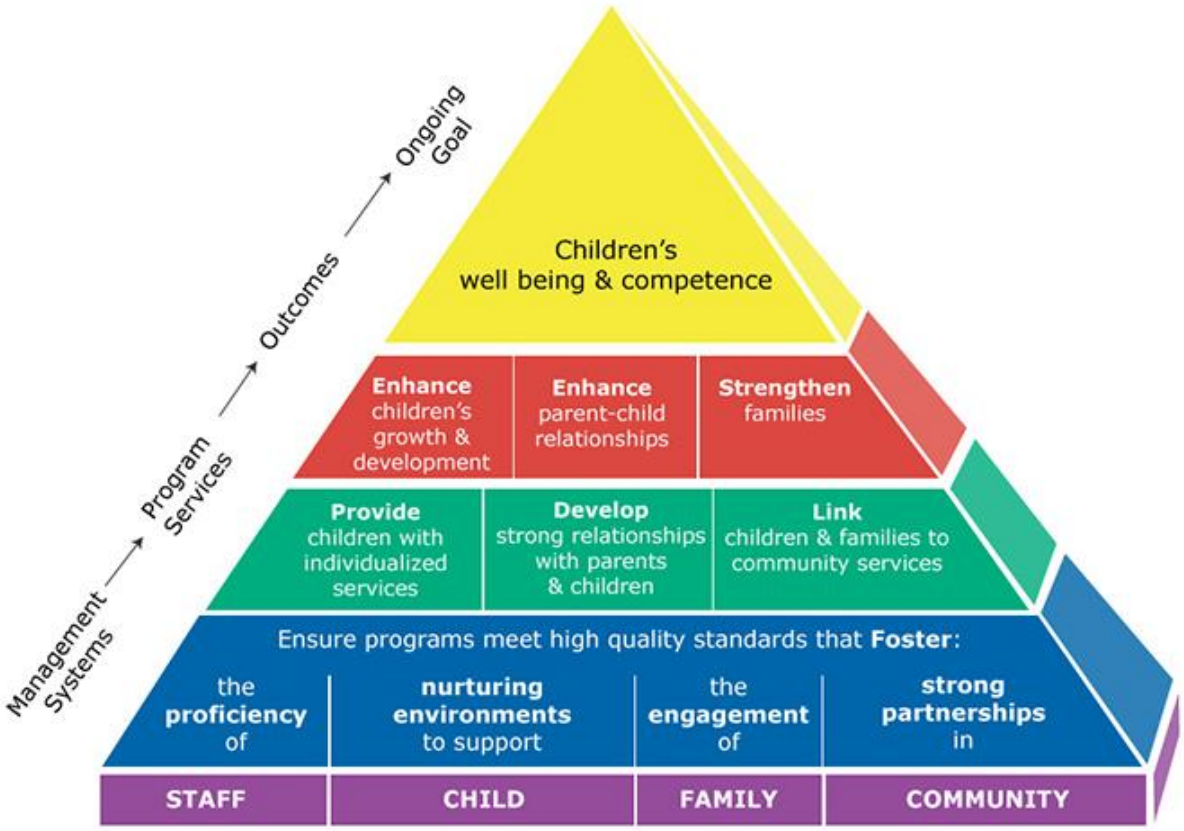
The Early Head Start program has served as a national laboratory by incorporating ongoing research to inform program improvement, which has evolved along with the program. The initial research on Early Head Start began concurrently with the launch of the program in 1995 (ACYF 2001; ACF 2002a; 2002b), and Baby FACES is the latest in a series of subsequent descriptive studies that have informed the program's development (see Vogel et al. 2006).

The Baby FACES Study

Baby FACES is charged with providing a national picture of the Early Head Start programs, services, staff, families, and children to inform program planning at national and local levels. The study was designed as a longitudinal descriptive study that includes a nationally representative sample of 89 programs, and children in a Newborn Cohort and 1-year-old Cohort through their experiences in the program. (Chapter II provides a more detailed presentation of the study design and sample.) For this final report, we include data gathered over children's entire experience in the program (or until they reached age 3). Because children were different ages in a given year, for the 1-year-old Cohort we collected age 3 data in spring 2011 and for the Newborn Cohort we collected these data in spring 2012. Similar to earlier rounds of data collection, we gathered information on participant families from parent interviews; staff (home visitors or teachers) reports on children; individual interviews with those teachers and home visitors about their characteristics, experience, and training; observations of study children's classrooms and home visits, weekly reports of services received by families (Family Services Tracking reports) and direct child assessments and video-recorded adult-child interactions conducted in the home.[2]

[2] In spring 2011 we conducted interviews with program directors, but did not collect program director interviews in spring 2012.

Figure I.1. Framework for Programs Serving Infants and Toddlers and Their Families



Source: Framework for Programs Serving Infants and Toddlers and Their Families. Head Start Approach to School Readiness. HHS/ACF/OHS. 2012.

The focus of this report is to describe how children and families are faring over time, the amount and quality of services they receive, and how child/family, staff, and program characteristics intersect and help to explain child and family outcomes at the end of the program. This report capitalizes on having longitudinal information on children and families and rich data on their experiences in the program. We describe children at age 3 and use available information to try to understand how particular characteristics of children, families, staff, and programs are related to given outcomes. As the only nationally representative study of Early Head Start to date, the findings presented here are our best understanding of what the program looks like. These analyses are exploratory and do not imply that findings are causal, but they may provide clues and avenues for continuous program improvement.

Chapter II details our approaches to answering the study's research questions. Box I.1 highlights key findings at age 2.

Box 1.1. Key Age 2 Findings

The age 2 report provided a snapshot of services offered and received, service quality and predictors of quality, children's development based on direct assessments of the 1-year-old Cohort as of spring 2010 when they were 2 years old, and an initial picture of those who left the program early (Vogel et al. 2015). Included here are highlights in each area.

Key Features of Services Offered and Received

Early Head Start programs offer core child development services at recommended frequencies. Based on data from the Family Services Tracking system, children in the home-based option were offered about one home visit per week for most of the year, with slightly fewer in summer and winter. Children in the center-based option were offered four to five days for most of the year; in summer and winter, the average number of days offered dropped to about three or four days per week.

Most families take up services at high frequencies, but rates vary by time of year. The average child in the home-based service option received 37 home visits per year, and the average child in the center-based option attended 179 days per year. These rates varied somewhat by child and family characteristics.

Key Features of Service Quality and Predictors of Quality

Observations of program quality, both in terms of center- and home-based services, placed programs in the mid-range, on average. Overall, classrooms score in the mid-range on the CLASS-T (between 3 to 5 out of a possible score of 7). Scores are highest in the area of Emotional and Behavioral Support. Scores are lowest in the area of Engaged Support for Learning. Aspects of classroom quality are positively related to teacher job satisfaction, experience, and child development credentialing, and negatively related to teacher depressive symptoms, staff turnover, and unfilled positions. HOVRS-A scores are highest (4 or higher out of 5) in the areas of Child Engagement and Relationship with the Family and lowest (less than 3) in Nonintrusiveness and Facilitation of Parent-Child Interaction. Scores in the area of Visitor Effectiveness are somewhat higher than those for Visitor Strategies. Home visit quality is positively related to home visitors' receipt of a CDA credential. Quality is also negatively associated with the number of unfilled staff positions in the program and home visitors' risk of depression.

Key Family and Child Features

Children's home environments are emotionally supportive and cognitively stimulating. Scores on the Home Observation for Measurement of the Environment (HOME) inventory (Caldwell and Bradley 2003) average nearly 25 out of 30, suggesting that 2-year-olds live in home environments that have adequate emotional support and cognitive and language stimulation.

Recordings of parent-child interactions show mid- to high-range levels of positive parenting behaviors and low levels of negative ones. According to observed ratings on the 7-point Parent-Child Interaction Rating Scales, parents of 2-year-old children received average scores of 4 (out of 7) on sensitivity, positive regard, stimulation of cognitive development, and relationship quality. Parental sensitivity, positive regard, and relationship quality were highly intercorrelated, and were combined into a single composite score (synchronicity), with comparable average ratings of 4. Conversely, negative parenting behaviors during the play-based assessment averaged 3 for negative regard, 4 for intrusiveness, 3 for detachment, and 3 for dissolution of boundaries.

Children's language development at 2 years old lags behind national norms. Two-year-old Early Head Start children score more than half a standard deviation below the national norms[3] on the English Preschool Language Scale-4 (PLS-4) Auditory Comprehension scale and on the Spanish PLS-4 Auditory Comprehension scale (91 and 90, respectively).

Children's social-emotional development varies depending on reporter. On the Brief Infant Toddler Social Emotional Assessment (BITSEA), staff reports yield Problem Scale scores that are comparable to the national norms, while parent reports are higher than the national norms. Both parent and staff reports rated more children at risk on the Competence subscale than the national norms. Assessors rate more 2-year-old children as scoring below the cutoffs on the Bayley Behavior Rating Scale (BRS), compared with the national norms.

Key Features of Early Exiters

Overall, exit rates are high, with more than one-fifth of children leaving Early Head Start by age 2. Children in home- and center-based options exit at similar rates. Early exiters attended programs with characteristics similar to

[3] The standard scores for the national normative sample have a mean of 100 and a standard deviation of 15.

those attended by continuing participants. Exiters' programs, however, have better staff-child ratios. Early exiters and continuing participants experience similar levels of classroom and home visit quality. Staff members serving early exiters have similar levels of education, experience, and depressive symptoms as those serving continuing participants. Early exiters and continuing participants share similar program attendance rates and were similar developmentally at age 1. Continuing participants' staff members rate staff-parent relationships slightly but significantly better than exiters' providers.

Road Map to the Report

Each chapter of this report includes information on key areas of the Early Head Start program and its families, and report appendixes A through D provide additional information about the technical details of the study and analyses.

- Chapter II includes information about study design and methods.
- Chapter III describes how the 3-year-olds who have remained in the program are faring.
- Chapter IV illustrates the developmental trajectories of children in Early Head Start using growth curve modeling that takes outcomes at earlier points in time into account.
- Chapter V describes observed quality of both home visits and classrooms and program implementation.
- Chapter VI reports on predictors of home visit and classroom quality.
- Chapter VII describes the services that programs offer and that families receive and sets the stage for modeling in Chapters VIII and IX.
- Chapter VIII reports on using child and family and program characteristics to predict aspects of service receipt/participation by families.
- Chapter IX builds on the work in earlier chapters to predict child and family outcomes from program experiences and quality.
- Chapter X discusses key themes and next steps for future research.

II. BABY FACES METHODOLOGICAL AND ANALYTIC APPROACHES

This chapter describes the study; outlines the research questions; and then describes our sample, measures, and approaches to analyses. Additional details about data collection procedures, assessor training, and analytic issues and our approaches to resolving them can be found in the appendices in Volume II. For more specific and detailed information about the study design, see our first report and technical appendices (Vogel et al. 2011).

Expert Input Informed Analytic Approaches

The Baby FACES team maintained its close collaboration with multiple stakeholders to continue shaping the study to be innovative and responsive to the policy, program, and research communities. As described in the baseline and first follow-up reports (Vogel et al. 2011, Vogel et al. 2015), we continued to confer with the study's technical working group (TWG) and held a meeting in October 2011 to present age 2 findings and solicit input for analysis of the current wave of data. In addition to the formal meeting, we solicited input on our analysis plan from various TWG members and outside experts. The study has benefited greatly from the exchange of ideas with and feedback from these experts.

As described in Chapter I, OPRE and OHS will use information from Baby FACES to inform Early Head Start program planning at the national and local levels. For the first time, we are able to answer questions about how children are faring over time, predictors of services received, and how these services (along with other characteristics) are related to outcomes.

This final report focuses on research questions that describe children who are 3 years old in 2011 (the 1-year-old Cohort) and in 2012 (the Newborn Cohort). In addition to answering several of the questions addressed in the previous report, this report answers new questions, some of which can only be answered now that multiple time points of data are available. Box II.1 outlines the research questions addressed in this and the two previous reports (Vogel et al. 2011, 2015).

Continued Focus on Comprehensive Data Collection from Multiple Reporters

Here, we provide a brief overview of the data elements collected in the final data collection periods (see Box II.2) and the measures used (Table II.1). Attributes of specific measures are detailed in the chapters in which they are reported and in the Technical Appendix.

Box II.1. Baby FACES Research Questions

Describing Early Head Start and Program Services (*Learning as We Go, Volume I*; *Toddlers in Early Head Start, Volume I*; current report Chapters V and VI)

- What is Early Head Start? What are the program models employed, the qualifications of staff, and other important program features and characteristics?
- What is the overall status of program implementation and quality?[1]
- What specific services are delivered to families?

Describing the Population Served (*Learning as We Go, Volume I*; *Toddlers in Early Head Start, Volume I*; current report Chapter IV)

- What are the characteristics of the families Early Head Start serves (includes demographic, household, and family characteristics; family needs; and risk factors)?
- How are Early Head Start children and families faring over time?

Relating Program Services to Child and Family Outcomes (current report Chapters VIII, and IX)

- How are child and family needs and outcomes associated with services received over time? Are there relationships between program features and outcomes?
- What are the characteristics of special populations and subgroups? What are the services provided for them? Examples of subgroups include children with identified special needs, highest-risk families, mothers with depression, DLLs, and mothers pregnant at program enrollment.[2]
- What family and child characteristics are linked to services received? What characteristics are linked to outcomes?

[1] The data collection approach in this study requires that we present the program quality findings at the child level because we did not sample centers or classrooms but instead observed study children in the setting they were receiving services. As described further in Chapter V, we can only make statements about the quality of care received by children in the study, and not draw conclusions about the quality of care in programs overall. The study children and the quality of the care they experience are representative of children of their respective ages who were enrolled in Early Head Start in 2009 and remained enrolled through age 3.

[2] The subgroups it is possible to analyze are limited by our sample size.

Box II.2. Overview of Baby FACES Data Sources and Measurement Approach at Age 3

Parent Interview. This interview asked parents (defined as the person primarily responsible for the care of the study child; how this child is identified is described below) about the demographic characteristics of the family and child, about their service needs and use, and about their well-being and that of the child. It also asked about the child's exposure to environmental health risks, and about environmental and routine supports for the child's growth and development. Parents were also asked to rate their child's development and behavior on a few assessments of their child's development. The interview was conducted by telephone. Study children were identified based on their age during spring 2009 when study enrollment occurred. The Newborn Cohort consisted of pregnant women and infants up to 8 weeks of age. The 1-year-old Cohort included children who were 10 to 15 months of age during the enrollment period in spring 2009.

Direct Child Assessment and Home Observation. Mathematica field staff conducted the child assessment home visits. The assessments include administration of the Preschool Language Scale-4 Auditory Comprehension subscale (PLS-4; Zimmerman et al. 2002), the Peabody Picture Vocabulary Scale–Fourth Edition (PPVT-4; Dunn and Dunn 2007) and measurement of height and weight. While in the home, the field assessor also observed the child's ability to focus on the tasks provided, the interactions between the child and parents, and the quality of the home environment as supports for children's safety and development (both internal and external). These observation measures include the Bayley Behavior Rating Scale (BRS; Bayley 2006), the Home Observation for Measurement of the Environment (HOME; Caldwell and Bradley 2003), and scales drawn from the Project on Human Development in Chicago Neighborhood (PHDCN, Sampson 2012).

Parent Self-Administered Questionnaire. While Mathematica field staff conducted the child assessment in the home, parents were asked to rate their child's development and behavior using the Ages & Stages Questionnaires, Third Edition (ASQ-3; Squires et al. 2009), MacArthur-Bates Communicative Development Inventories (CDI; Fenson et al. 2000), the Brief Infant Toddler Social Emotional Assessment (BITSEA; Briggs-Gowan and Carter 2006), and the Behavior Problems Index (BPI; Zill and Peterson 1986). They also rated the quality of their relationship with the child's home visitor or teacher.

Parent-Child and Assessor-Child Interaction. Children participated in two semi-structured interaction activities, playing with two sets of toys. These interactions were video-recorded for later coding of target behaviors. First, the parent and child were asked to sit on a mat and play with the contents of two bags of toys. This activity is known as the Two Bags task, an adaptation of the parent-child interaction task used in the EHSREP and the Early Childhood Longitudinal Study-Birth Cohort (ECLS-B). Second, the field assessor and the child played with another set of toys, with the assessor following a standardized protocol (the Early Communication Indicator [ECI; Greenwood et al. 2006]).

Staff-Child Report. Home visitors and teachers of study children completed a child-specific rating with details on the characteristics of the families and children, such as rating a family's participation over the past 6 months, children's vocabulary on the CDI, and their behavior on the BITSEA and BPI. Staff members also rated the quality of their relationship with the parents of study children.

Classroom Quality Observation. To assess the quality of center-based services that study children received, a Mathematica field staff member observed the quality of children's classrooms using a set of measures. These measures included the number of children in the classroom and number of adults caring for them, as well as the quality of the materials and the interactions between children and their teachers. For 3-year-olds, observers used the Classroom Assessment Scoring System-Toddler version (CLASS-T; Pianta et al. 2010). It is important to note that these observations assess only the quality of care received by the study children and are not necessarily representative of the quality of care received by all children in the program. As described below, we sampled programs and attempted to recruit into the study all children in our age-eligibility windows; we did not sample at the classroom or home visitor level. This approach limits the ability to generalize from the findings to the quality of care in the Early Head Start program overall.[1] If more than one study child was in the same classroom, we conducted only one observation.

Home Visitor/Teacher Interview. We interviewed children's home visitors and teachers to determine their demographic characteristics, tenure working for the program, and well-being, as well as training and

educational experiences provided by the program and the working environment. We interviewed only staff members working with study children at the time of each data-collection wave and did not follow staff members over time unless they were still working with one or more study children at multiple waves.

Home Visit Quality Observation. To assess the quality of the home visits that study children received, field assessors observed the home visitors who provided services to children in the study sample using the Home Visit Rating Scales-Adapted (HOVRS-A; Roggman et al. 2009) and a form that assessed the content and characteristics of the visit. The home visit observations had the same generalizability limitations as the classroom observations just described. We scheduled an observation of each home visitor who had a study child on his or her caseload, but did not observe a home visit for each child in the home visiting option.[2]

Family Services Tracking (FST). To capture the services received by families, Early Head Start home visitors and teachers of study children completed a weekly service tracking form that detailed the number of service experiences (home visits or days in care) study children were scheduled to have and the number that they actually received.

[1]The data collection approach requires that we present the program quality findings at the child level. As described further in Chapter V, we can make statements only about the quality of care received by children in the study.

[2]Before deciding on this approach (observing one visit per home visitor rather than one visit per child), we consulted with Lori Roggman, Ph.D., one of the developers of the home visit observation tool we used, to better understand how much variability within a home visitor we might expect. According to Roggman, home visits tend not to vary much in the observed quality ratings over different families (personal communication, April 2009) for the same home visitor. That is, home visitors' adeptness at conducting a home visit seems to translate into a quality level independent of the family they are visiting. Given the enormous constraints in scheduling observations of visits during the brief time we had on site, we opted to observe each home visitor only once and apply those ratings to all the study children on that home visitor's caseload.

Table II.1. Key Measures Used in This Report

Program Characteristics and Implementation	
Program Approach: Program Level	Program approach at the program level is based on director responses to questions regarding (a) the types of services their programs offer (center-based, home-based, or combination) and, (b) separately for each service option, the frequency of services offered, using responses from 2009, 2010 and 2011.
Program Option: Family Level	Program approach at the family level is based on information collected during interviews with parents and with information from programs in our Sample Management System (SMS). Parents were asked whether they receive center-based services, home-based services, family child care services, or another type of service (such as a combination of services). Parents also indicated the frequency of center attendance and home visits received.
Program Implementation: Program Level	Program directors completed a self-rating of their implementation in four cornerstone areas in an SAQ in 2009. In 2010 and 2011, they answered questions in the Program Director Survey that were then scored by the analysis team. For this report, we calculate cross-year averages of the 2009-2011 cornerstone ratings. Programs with cross-year ratings equal to or above 3 on each of the cornerstones are designated as "fully implemented."
Population Served	Program directors reported the proportion of families enrolled in their programs who were facing socio-demographic and psychological risks. We use information from 2009 and 2010 to identify programs serving a high proportion (50 percent or more) of families facing these risks.
Staff Characteristics and Program Quality	
The Center for Epidemiologic Studies Depression Scale—Short Form (CESD-SF; Radloff 1977; Ross et al. 1983)	The CESD-SF is the short form of the full-version CESD, which is a self-administered screening tool used to identify symptoms of depression or psychological distress. The tool was used to measure depression symptoms in teachers and home visitors.
Parent-Caregiver Relationship Scale (PCRS; Elicker et al. 1997)	The PCRS was reported by staff and measures the perceived relationship between the parent and the teacher or home visitor of infants and toddlers. Items capture important dimensions of the parent-caregiver relationship, including trust and confidence, communication, respect/ acceptance, caring, competence/ knowledge, partnership/ collaboration, and shared values.
Staff Demographic Characteristics	The teacher and home visitor interviews included sections with items that broadly covered: parent participation in the program, staff training and supervision, staff benefits and morale, languages spoken (by the staff member and by families in the classroom or caseload), racial/ethnic group membership, and education.
Home Visit Rating Scale-Adapted (HOVRS-A; Roggman et al. 2009), modified from the HOVRS (Roggman, Cook, Jump, Boyce and Innocenti 2006b)	Observations of home visits used the HOVRS-A, an adaptation of the HOVRS (Roggman et al. 2006b). The HOVRS-A consists of 7 items measuring the quality of home visitor strategies and effectiveness at involving and engaging the family during home visits.
Home Visit Characteristics and Content (Boller et al. 2009)	During structured observations of home visits, field staff also collected data on the topics covered, activities, and structure of the home visit.
Classroom Assessment Scoring System, Toddler Version (Pianta, La Paro, and Hamre 2010)	The CLASS-T (Pianta et al. 2010) was used for classroom observation. It is an adaptation of the Pre-K CLASS (Pianta et al. 2008), which focuses on teacher-child interaction quality in toddler child care classrooms. The CLASS-T measures process quality along eight dimensions (Positive Climate, Negative Climate, Teacher Sensitivity, Regard for Child Perspectives, Behavior Guidance, Facilitation of Learning and Development, Quality of Feedback, and Language Modeling) within 2 domains: Emotional and Behavioral Support and Engaged Support for Learning. Dimensions are defined by observable indicators along a 7-point scale, with ratings reflecting scores in the low (1-2), mid (3-5), and high (6-7) ranges.
Child-Adult Ratio	Center-based classroom observations also included child-adult ratios and group sizes.

Services Offered and Received	
Length of program enrollment	Length of enrollment is a continuous variable expressing time of enrollment in Early Head Start in months between initial date of enrollment and graduation or drop out.
Family involvement in past 6 months	Ratings of involvement of the family by staff and collected at each wave of data collection in the Staff-Child Report (SCR). Families are rated on a 4-point scale from not involved to consistently highly involved.
Services offered by programs	This reflects home visits or center days offered to families during their enrollment. The information is derived from weekly Family Services Tracking (FST) data completed by the child's teacher or home visitor on "regularly scheduled" home visits and center days, and adjusted for missed services due to program reasons.
Services received by families	This reflects home visits or center days received during their enrollment. The information is derived from weekly staff-reported FST data.
Family Characteristics, Parenting, and the Home Environment	
Maternal Demographic Risk Index (ACF 2001)	The maternal demographic risk index captures the multiple dimensions of risk of poor developmental outcomes a child may face as a consequence of his or her mother's socioeconomic circumstances. The index comprises three risk groups (lower, medium, and highest). The index was constructed by summing the number of the following risk factors that the mother reported she faced: (1) being a teenage mother, (2) having no high school credential, (3) receiving public assistance, (4) not being employed or in school or training, and (5) being a single mother.
Maternal Psychological Risk (ACF 2001)	This is an index of cumulative risk based on mothers' reports of (1) moderate or severe depressive symptoms, (2) parenting stress one standard deviation or higher than the sample mean on either the Parenting Stress subscale or the Parent-Child Dysfunctional Interaction subscale of the Parenting Stress Index, and (3) substance use problems including parent reports of drug use in the past year or having ever had a drug or drinking problem. Scores are classified as no risk (0 risk factors), medium (1 risk factor), and high (2 or 3 risk factors).
Maternal and Child Characteristics	The parent interview also included sections that broadly covered many different aspects of the family and home environment, including family racial/ethnic membership, language(s) spoken in the home, program services received, parent and child health, family routines, income and housing, and income and needs.
Parent-Child Interaction Rating Scales for the Two-Bag Assessment (PCI): Parenting Behaviors (Mathematica Policy Research 2010	Semi-structured, video-recorded assessments of parent-child play (Two-Bag Task) were coded using the Parent-Child Interaction (PCI) Rating Scales for the Two-Bag Task. The PCI consist of 12 scales that assess a range of child and parent behaviors. Each of eight parent behaviors is rated along a 7-point scale, ranging from a very low incidence of the behavior to a very high incidence of the behavior. A composite parenting score, synchronicity, was derived by averaging scores on parental sensitivity, positive regard, and relationship quality—all of which were highly and significantly correlated.
The Parenting Interactions with Children: Checklist of Observations Linked to Outcomes (PICCOLO; Cook & Roggman 2009; Roggman et al. 2009)	The PICCOLO is an observational instrument designed to measure positive parenting along four domains known to support children's early development: affection, responsiveness, encouragement, and teaching. Twenty-nine behaviors are rated on a 3-point scale, ranging from 0 (absent) to 2 (clearly evident). Behaviors that are infrequently observed are indicated by a score of 1. The domains of affection, responsiveness, and encouragement each consist of seven items; the teaching scale consists of eight items.
Home Observation for Measurement of the Environment (HOME; Caldwell and Bradley 1984)	The HOME measures the quality of stimulation and support available to a child in the home environment. Information needed to score the inventory is obtained through a combination of parent self-report and assessor observation conducted in the home with the child's parent while the child is present. We used selected items from the Infant version of the HOME inventory, the internal environment items from the Early Childhood version of the HOME, and neighborhood rating items from the PHDCN (Sampson 2012). We derived five subscales from this assessment, as well as the total score.

Parent Support for Child Learning Index	This composite index captures the degree of cognitive stimulation provided to children in the home setting by parents or other household members. Component measures include PCI Cognitive Stimulation scores; PICCOLO Teaching scale scores; verbal responsiveness items from the HOME (whether the parent converses with the child at least twice during the visit, answers the child's questions/requests, responds to the child's talk verbally, and uses complex sentence structure in social exchanges); parent-reported frequency of bookreading and storytelling by any household member; and whether any household member helped the child learn shapes/sizes, the alphabet, colors or numbers. Scores were derived by averaging z-scored values for each of the components
Parent Provision of Learning Materials Index	The provision of home learning materials was assessed through a combination of interview items and interviewer observations from the HOME scale. Component items include the number of accessible children's books in the home; the availability of puzzles, toys that teach colors/sizes/shapes, and toys that teach numbers; the accessibility of toys, games, and books appropriate for preschoolers; and whether the child had access to an audio device and at least 5 children's tapes. Scores were derived by averaging z-scored values for each of the components.
Exposure to Violence	Exposure to Violence measures how many violent incidents (out of four) a child has observed or been a victim of in his or her lifetime, according to parent reports. Items come from the Infant-Toddler Social and Emotional Assessment (Carter and Briggs-Gowan 2000), in which parents are asked to respond yes or no to questions such as whether a child has "seen violence in their neighborhood" or "seen someone hit, push or kick a family member."
External Environment	Using items from the PHDCN, external environment is a measure of the physical and social environment of the face-block (roughly equivalent to the street between two cross streets, or about 10 housing units) where the family lives. Items in this subscale are based entirely on assessor observations of the neighborhood, and include such items as general condition of most of the housing units, garbage in the street or on the sidewalk, volume of traffic, and people arguing or fighting in the street. The items are recoded as 1 (yes) or 0 (no), and then summed. Scores can range from 0 to 8.
Neighborhood Disorder	This construct uses the same items as the External Environment construct above, but scored as a z-score. The scale score is the mean of the item z-scores. Higher scores indicate higher levels of disorder.
The Center for Epidemiologic Studies Depression Scale—Short Form (CESD-SF; Radloff 1977; Ross et al. 1983)	The CESD-SF is the short form of the full-version CESD, which is a self-administered screening tool used to identify symptoms of depression or psychological distress. The tool was used in Baby FACES to measure depressive symptoms of mothers (as well as in care providers).
The Parenting Stress Index—Short Form (PSI-SF; Abidin 1995)	The PSI-SF measures the degree of stress in parent-child relationships. Baby FACES includes two subscales: (1) the Parental Distress subscale measures the level of distress the parent is feeling in his or her role as a parent; and (2) the Parent-Child Dysfunctional Interaction subscale measures the parent's perception that the child does not meet expectations and that interactions with the child do not reinforce the parent.
The Family Environment Scale, Family Conflict Subscale (FES; Moos 2002)	The FES was designed to measure the social and environmental characteristics of families. The Family Conflict subscale measures the extent to which the open expression of anger and aggression and conflict-filled interactions are characteristic of the family.
Child Development	
Preschool Language Scale—Fourth Edition (PLS-4; Zimmerman et al. 2002a, 2002b).	The PLS-4 is a direct child assessment used to evaluate receptive and expressive language skills, as well as understanding and use of grammatical rules for children from birth to 6 years of age. It is composed of two subscales: Auditory Comprehension (AC) and Expressive Communication (EC). We used the AC subscale for both of the English and Spanish editions of the PLS-4.
Early Communication Indicator (ECI; Luze et al. 2001; Carta et al. 2010).	The ECI is a semi-structured, play-based assessment designed to measure the expressive communication of infants and toddlers between the ages of 6 and 36 months along four key skill elements: gestures, vocalizations, single-word utterances, and multiple-word utterances. Assessors administered the ECI which was video recorded for later coding by staff at Mathematica.

Peabody Picture Vocabulary Test-4th Edition (PPVT-4;Dunn and Dunn 2007)	The PPVT-4 is a measure of receptive vocabulary in which children are shown a plate with four pictures and asked to point to the one that indicates the target word that is stated by the assessor "point to [target word]." It is a norm-referenced standardized test and is suitable for a wide range of ages, from 2½ through adulthood.
MacArthur-Bates Communicative Development Inventories—Infant Short Form (CDI; Fenson et al. 2000)	The CDI is designed to assess children's early receptive and expressive language and communication skills through parent report. Two measures were derived from this form: vocabulary comprehension and vocabulary production.
Ages & Stages Questionnaires, Third Edition (ASQ-3; Squires et al. 2009)	The ASQ-3 is a parent-report tool for screening children from 1 month through 5-1/2 years of age for developmental delays in five key developmental areas: (1) communication, (2) gross motor, (3) fine motor, (4) personal-social, and (5) problem solving.
The Brief Infant Toddler Social Emotional Assessment (BITSEA; Briggs-Gowan and Carter 2006)	The BITSEA is the screener version of the longer ITSEA, which is designed to detect delays in the acquisition of social-emotional competencies as well as social-emotional and behavior problems in children 12 to 36 months old.
Parent-Child Interaction Rating Scales for the Two-Bag Assessment: Child Behaviors (Mathematica Policy Research 2010)	Semi-structured, video-recorded assessments of parent-child play (Two-Bag Task) were coded using the Parent-Child Interaction (PCI) Rating Scales for the Two-Bag Task. The PCI includes four child behaviors rated along a 7-point scale, ranging from a very low incidence of the behavior to a very high incidence of the behavior.
Bayley Behavioral Rating Scale (BRS; Bayley 1993)	The BRS measures the child's behavior during child assessment. The BRS is one of the three component scales of the Bayley Scales of Infant Development—Second Edition (Bayley 1993). There are two subscales of the BRS used in Baby FACES: Orientation/Engagement, measuring the child's cooperation with the assessor during the assessment, positive affect, and interest in the test materials, and Emotional Regulation, measuring the child's ability to change tasks and test materials, negative affect, and frustration with tasks during the assessment.
Behavior Problems Index (BPI; Zill and Peterson 1986)	The Behavior Problems Index (BPI) measures child externalizing behavior problems (such as aggression and hyperactivity) and internalizing behavior problems (such as anxiety and depression). The NLSY used the BPI with children 4 years of age or older and the PSID Child Development Supplement used the BPI with children 3 years of age or older. Baby FACES used the BPI in Staff Child Report and Parent Self-Administered Questionnaire, but the items are slightly different than those used in these two studies.

Note: Each chapter presents additional information about the measures used in it. Appendix C describes the psychometric properties of constructed variables.

Response Rates Are High Across Study Waves

Maintaining high response rates in a longitudinal study is difficult, but the study achieved high rates of completion for each of the data collection instruments over time. Table II.2 summarizes response rates for the entire study as a percentage of the respondents or settings eligible to participate in each collection. (See Appendix B for a breakdown by cohort.) Table II.3 describes the sample at each wave of data collection in relation to the baseline sample.

Family Services Tracking Reports on Weekly Service Receipt

The data we analyze for this report covers a 104-week period from the spring of each child's age 1 year to the spring of his/her age 3 year.[4] During this period we received reports for 793 children in total. From age 1 to 2, we received at least one FST report for 772 children from staff. From age 2 to 3, staff completed at least one report for 517 children. The sample is smaller in the age 2 to 3 year in part due to program exits (170 children) but also due to staff nonresponse. A child who did not leave

[4] This period spans July 2009 to June 2011 for the 1-year-old Cohort and July 2010 to June 2012 for the Newborn Cohort.

his or her Early Head Start program early was eligible for 52 reports in each year. On average, from age 1 to age 2, reports were submitted for 73 percent of eligible weeks. From age 2 to 3, reports were submitted for 69 percent of eligible weeks.[5]

Table II.2. Baby FACES Response Rates 2009 – 2012

Instrument	2009 Number Completed (Percentage)	2010 Number Completed (Percentage)	2011 Number Completed (Percentage)	2012 Number Completed (Percentage)
Staff-Child Report	933 (95.5)	703 (95.6)	538 (96.2)	82 (97.6)
Parent Interview (CATI)	894 (91.7)	583 (79.3)	445 (79.6)	61(72.6)
Parent Self-Administered Questionnaire (SAQ)	n.a.	537 (89.5)	481 (86.0)	70 (83.3)
Child Assessment	n.a.	547 (91.2)	503 (90.0)	76 (90.5)
Caregiver Interview	229 (93.1)	267 (98.9)	232 (98.7)	44 (100)
Home Visitor Interview	323 (96.7)	225 (97.0)	174 (99.4)	29 (100)
ITERS-R	223 (94.9)	53 (98.1)	n.a.	n.a.
CLASS-T[a]	n.a.	220 (98.7)	231 (99.1)	42 (95.5)
HOVRS-A	242 (89.3)	193 (83.2)	139 (84.2)	20 (87.0)
Program Director Interview	89 (100)	89 (100)	89 (100)	n.a.
Program Director SAQ	86 (96.6)	83 (93.3)	n.a.	n.a.
Exit Interview[b]	62 (54.9)	76 (38.3)	98 (57.9)	337 (72.8)

Source: Baby FACES Sample Management System (SMS).

Note: Percentages are of those still enrolled in Early Head Start and therefore still eligible for the study at each wave. Exit interviews are of those who were reported to have left the program at each time period.

[a]Administered in classrooms of 2- and 3-year-olds. The ITERS-R was used to rate classrooms of 1-year-olds.

[b]Administered to parents of children who left the program by each round of data collection as well as those who graduated from the program. Round 1 was administered between October and December 2009. Round 2 was administered between April and June 2010. Round 3 was administered February to June 2011. The total number of cases released for Round 2 includes the 51 incomplete responses from Round 1. Overall, between Rounds 1 and 2, we released a total of 258 unique cases. The combined response rate of Rounds 1 and 2 is 54 percent. The final round was administered to parents of 1-year-old Cohort children who remained enrolled through age 3, when the children were 3 1/2 years old (August through October 2011).

ITERS-R = Infant Toddler Environment Rating Scale-Revised; HOVRS-A = Home Visitor Rating Scale-Adapted; CLASS-T=Classroom Assessment Scoring System-Toddler version; n.a. = not applicable.

[5] Staff members could enter FST reports directly into our web-based system or complete paper forms we supplied and mail them for us to input on their behalf.

Table II.3. Baby FACES Response Rates 2009—2012 As a Percentage of the Baseline Sample

Instrument	2009 Number Completed (Percentage)	2010 Number Completed (Percentage)	2011 Number Completed (Percentage)	2012 Number Completed (Percentage)
Staff-Child Report	933 (95.5)	703 (72.0)	538 (55.1)	82 (42.3)
Parent Interview (CATI)	894 (91.7)	583 (59.7)	445 (45.6)	61 (31.4)
Parent Self-Administered Questionnaire (SAQ)	–	537 (68.9) [a]	481 (49.3)	70 (36.1)
Child Assessment	–	547 (70.2) [a]	503 (51.5)	76 (39.2)

Source: Sample Management System.

Note: Percentages are of those eligible and consenting to participate in Baby FACES in spring 2009.

[a]Administered to 1-year-old Cohort children and parents.

Attrition and Comparison of Graduates and Early Exiters

Only families who are enrolled in the Early Head Start program are eligible for Baby FACES at each wave of data collection. Attrition from the program is a factor in our analyses. We consider those who stayed until age 3 as "graduates" of the program. In all, 253 families out of 782 originally enrolled in the 1-year-old Cohort left Early Head Start before their child was 3 years old,[6] and 108 families out of 194 originally enrolled in the Newborn Cohort did so.

Differences between graduates and early exiters can introduce bias in our estimates of age 3 outcomes. For example, if children from more disadvantaged backgrounds are more likely to exit from the program, a comparison of average outcomes at baseline and age 3 may lead to overestimation of the gains children make. The presence of exiters may pull down the baseline mean and their absence from subsequent assessment may make the mean in later years look higher. Conversely, if we are interested in describing the outcomes of families who enroll and stay for the full term of the program, it may not be prudent to include early exiters in our estimates at all. In Appendix D we provide extensive discussion of the baseline differences and similarities between the characteristics of graduates as compared with early exiters.

In our analysis of children's growth over time (Chapter IV), we fit our models both with and without the early exiters in the sample to provide a sense of how results differ. It is important to note, however, that our analysis of stayers may not be generalizable to exiters. That is, we cannot assume that the outcomes we observe for stayers are the same that we would expect to see among exiters if they had stayed in the program.

The longitudinal nature of Baby FACES adds complexity to describing the sample at each time point and overall. A description of the sample needs to include both point-in-time and longitudinal estimates so that sample selection issues are transparent. In Table II.4 we present descriptive statistics for the entire sample of 3-year-olds (combining cohorts). In prior reports we have presented information separately for the Newborn and 1-year-old Cohorts to better understand how they are similar and different. Differences between the cohorts initially were in expected directions (for instance in terms of maternal employment—lower in the Newborn Cohort and in terms of depression—higher in the 1-year-old Cohort). By the time of the first follow-up when the Newborns were 1 year old, they were more similar to the older cohort at that age. Because the Newborn Cohort is very small we now combine cohorts for analyses to increase sample size.

[6] We defined the lower bound of age 3 as 32 months of age since transitions are based in part on age and in part on other factors, such as the time of year and availability of Head Start slots, and it is possible for children to graduate prior to 36 months of age.

Table II.4. Demographic and Household Characteristics of the Baby FACES Sample Over Time (Percentages Unless Otherwise Indicated)

	All 1-year-olds (Standard Error)	All 2-year-olds (Standard Error)	All 3-year-olds (Standard Error)
Child's Race/Ethnicity			
Hispanic[a]	35.3 (3.58)	35.5 (3.96)	36.2 (4.18)
White, non-Hispanic	36.2 (3.51)	36.3 (3.82)	35.6 (3.99)
African American, non-Hispanic	17.8 (2.96)	18.4 (3.20)	18.4 (3.47)
Multiracial, non-Hispanic	7.9 (1.43)	7.1 (1.49)	6.9 (1.70)
American Indian/Alaska Native	1.6 (0.67)	1.3 (0.52)	1.4 (0.66)
Other	1.3 (0.47)	1.4 (0.56)	1.4 (0.61)
Parent Immigrant Status			
Both parents born in U.S.	70.9 (3.19)	68.8 (3.49)	67.0 (3.76)
One parent born outside U.S.	9.8 (1.14)	9.5 (1.30)	9.3 (1.64)
Both parents born outside U.S.	19.3 (2.94)	21.7 (3.12)	23.7 (3.37)
DLL	35.9 (3.42)	35.8 (3.73)	35.9 (3.95)
Mother's Average Age in Years at First Child's Birth	20.4 (0.21)	20.5 (0.24)	20.6 (0.25)
Mother's Race			
White, non-Hispanic	32.1 (3.60)	32.3 (3.91)	33.1 (4.01)
Hispanic	44.0 (3.78)	43.7 (4.02)	42.9 (4.09)
African American, non-Hispanic	17.8 (3.00)	18.1 (3.17)	17.6 (3.31)
American Indian/Alaska Native	3.4 (0.82)	3.3 (0.96)	3.3 (0.97)
Multiracial, non-Hispanic	1.4 (0.71)	1.4 (0.67)	1.6 (0.87)
Other	1.3 (0.54)	1.2 (0.64)	1.6 (0.88)
Child Lives with			
Two biological parents	48.4 (2.74)	52.9 (2.97)	52.5 (3.25)
Married	59.5 (2.47)	59.4 (3.00)	64.0 (3.90)
Unmarried	40.5 (2.47)	40.6 (3.00)	36.0 (3.90)
One biological parent	46.6 (2.60)	37.5 (2.68)	35.4 (2.55)
No biological parents	1.9 (0.55)	3.0 (0.80)	2.6 (0.93)
Among Children Living Without Birth Father			
Child lives with Father Figure	11.6 (1.87)	16.7 (3.00)	17.5 (3.26)
Mean Number of Adults in Household	1.9 (0.04)	2.0 (0.05)	1.9 (0.06)
Mean Number of Children in Household	2.5 (0.06)	2.7 (0.09)	2.7 (0.08)
Average Household Size	4.4 (0.08)	4.7 (0.11)	4.6 (0.10)
Child Lives in Intergenerational Household	16.8 (1.84)	8.1 (1.37)	4.3 (1.16)
Average Household Income	24,880.4 (1973.65)	21,610.6 (696.72)	20,249.6 (723.53)
Median Household Income	17,500.0	18,000.0	18,000.0
Household Income			
$0–$9,999	24.2 (1.68)	21.1 (2.14)	19.3 (2.18)
$10,000–$17,499	23.5 (1.69)	24.4 (2.22)	25.9 (2.39)
$17,500–$24,999	20.8 (1.99)	23.1 (2.68)	26.5 (2.55)
$25,000 or more	31.4 (1.74)	31.3 (2.21)	28.2 (2.47)
Household Income as a Percentage of the Poverty Level[b]			
0–50	29.1 (1.83)	30.2 (2.55)	29.7 (2.54)
51–100	38.9 (2.22)	35.5 (2.60)	40.5 (2.90)
101–130	12.4 (1.11)	17.7 (1.89)	12.5 (2.09)
131–higher	19.6 (1.70)	16.5 (1.70)	17.3 (2.10)
Sample Size	**341-389**	**245-661**	**184-530**

Source: Spring 2009, 2010, 2011, and 2012 Parent Interview.

Note: Sample sizes reported at each point reflect the respondents to that wave of data collection. Cohorts are combined so that information on 1-year-olds includes information collected in 2009 for the 1-year-old Cohort and in 2010 for the Newborn Cohort, and so on for the other years.

[a] Throughout this report, Hispanic refers to those with a Hispanic or Latino racial/ethnic background.
[b] Poverty level is adjusted for household size according to the relevant year's HHS poverty guidelines.

DLL=Dual language learner

Data Collection Procedures and Field Training Were Rigorous

Our overall approach to training and data collection was rigorous and comprehensive. Training had three primary components: (1) observations of home visit and classroom quality (using the HOVRS-A and the CLASS-T), (2) conducting direct assessments of 3-year-olds, and (3) coding video-recorded parent-child and assessor-child interactions.[7]

Training and Certification for Observations

The procedures for the HOVRS-A and CLASS-T were similar to those employed in the prior round of data collection. (See Appendix B.) Again during the field period, quality assurance staff observed a home visit and classroom for each of the field staff to ensure that they continued to be reliable. Although rare in occurrence, if an observer was deemed no longer reliable, a team leader conducted observations and worked with the observer until he or she once again achieved reliability.

Training and Certification for Child Assessments

In the final rounds of data collection, field staff conducted direct assessments of 3-year-olds (the 1-year-old Cohort in 2011 and the Newborn Cohort in 2012). Training consisted of question-by-question explanations on the PLS-4 and the PPVT-4, instruction on the operation of the camera and tripod, and an explanation of the Two Bags and ECI video-recorded play tasks. Field staff had the opportunity to practice conducting these assessments with a 3-year-old child, with trained gold standards observing and rating their performance. After the training, assessors were certified to conduct the assessments. Ultimately, a group of 22 field staff was certified to conduct assessments in 2011 and 9 of the assessors returned in 2012. Quality assurance was accomplished by review of a video each staff member sent in of an early administration of the child assessments. Anyone not passing the first review received individualized feedback and needed to send in another video. (See Appendix B for details of the quality assurance process.)

Video Coder Training and Certification

As in the prior wave of data collection, teams at Mathematica coded the video assessments. In 2011 we used three teams, each coding one of the schemes used (parent-child interactions were coded with the Parent-Child Interaction Rating Scales (PCI) and the PICCOLO. The ECI was administered as the assessor-child play task. Coders for all schemes needed to achieve initial reliability with a gold standard and then maintain it in weekly reliability checks. (See Appendix B for details.) Because the sample was so small in 2012, we used former team leaders (three teams of two people each) to code all videos.

Approach to Weighting

Similar to the baseline report, this report presents descriptive information on program, families, and children in our sample. Because we selected a nationally representative sample of programs (and included all children meeting study criteria for our two cohorts), we use weights so that our findings generalize to all Early Head Start programs and children in our birthday windows who were enrolled in spring 2009 and continued to be enrolled at each data collection point. We constructed both cross-sectional and longitudinal weights. The weighting steps for each involved first defining the eligible

[7] Telephone interviewers completed the parent interview using computer-assisted telephone interviewing (CATI). The interviewers attended a two-day training session conducted by project staff on administration procedures, and Survey Operations Center supervisors monitored calls regularly.

population for that particular weight, adjusting for parental consent (and for some weights, adjusting for the presence of a parent interview in the same step), then adjusting for various combinations of completed instruments as defined for each weight. These combinations were designated based on analytical needs. We balanced several competing issues when determining how many, and which, weights to construct. Calculating weights for every variable and every combination of variables would have been an overwhelming task and too difficult to make use of during analysis, so we tried to be parsimonious in the number of weights constructed. Appendix A details the weighting procedures.

Child-Level Weights

Child-level weights are based on child-level data (such as staff child reports and child assessments) or staff-level data (such as the child's staff interviews and classroom or home observations). We first constructed a child-level weight to predict a "complete," which indicates both parental consent and (for many weights) at least one parent interview. By applying a nonresponse adjustment, we weighted up children with consent (and with a parent interview) to reflect all eligible children (by cohort and year of data collection). All children with a positive value for the weight requiring a parent interview had demographic data from that interview that could then be used for nonresponse adjustments in the child-level weights involving those other instruments.

Longitudinal Weights. We also constructed a set of longitudinal weights targeted for specific planned analyses of Baby FACES data over time. Each one requires eligibility over the appropriate age range, and parental consent as of the end of the study.

Exit Weights. To compare the characteristics of children and families who exited the Early Head Start program during the first study year with characteristics of those who remained, we created a child-level "consent" weight. This weight adjusted for the program's probability of selection and its participation, and whether the child had parental consent. We also generated a weight that accounts for whether the parent completed the exit interview, among those children who prematurely exited the program.

FST Weights. We constructed an FST weight for age 1 (meaning we had FST data between ages 1 and 2) and for age 2 (meaning we had FST data between ages 2 and 3), and one for when we had FST for either. We also constructed a weight for whether we had FST in year 1 (regardless of age). Again, we started with the consent-adjusted weight. Then, using the same stepwise procedures described above, we adjusted for whether we received any FST data for the child over the course of the time period defined for that weight.

Analytic Approaches and Limitations

In this final report we use data on children's entire experience in Early Head Start to examine longitudinal development and relationships among program and family characteristics, Early Head Start experiences (participation in and quality of services), and outcomes. In this section we provide some detail about how we approach problems of missing data and the potential for selection bias in our sample, and given those limitations, the methods we use to answer each of the research questions.

Early Head Start is a program that was created with the idea that participation in its services will lead to better outcomes for children and families. Program impacts on children and families have previously been documented elsewhere (ACF 2002a). Baby FACES was designed to extend findings from previous experimental research by delving more extensively into variation in program implementation and family participation. The motivation behind this line of inquiry is to better

understand whether and how services can be improved, targeted, and delivered to facilitate even better outcomes for more families. For example, prior research (ACF 2002b) found that families facing a high number of risk factors did not demonstrate benefits from the program. Baby FACES is uniquely positioned to examine why this might be the case. For example, are high-risk families receiving lower quality services? Do high-risk families receive the recommended amount of services?

Because Baby FACES is not an experimental study, there are certain limitations in our analysis and conclusions, many of which are the result of selection bias. Families have chosen the programs they are enrolled in and have made subsequent decisions about whether to stay enrolled in the program and how much to participate. Because we have detailed information about families from multiple sources, we can use those data to compare families within the study only to their similar counterparts. However, there is always a possibility that we have failed to account for some differences between families in characteristics that we did not observe. This is a limitation in a non-experimental study, where the services families receive are self-selected or determined by what the program offers. The results provide substantial detail of our analyses and the different variables we include in our models. We also offer discussion of any remaining potential threats to the validity of results, where applicable. Although we do our best to investigate and minimize bias in estimates by employing rigorous statistical methods, information on the limits of our design and of these non-experimental methods will allow readers to judge the extent to which our estimates reflect the causal effects of services on outcomes. More detailed information on statistical approaches is provided in Appendix D.

Descriptions of Children's Development at Age 3 and Growth Over Time

We use simple descriptive statistics such as means/percentages and standard errors (weighted as appropriate) to depict 3-year-old children and their families in our sample. As we have done in the prior two reports, where possible we compare our sample to a normative one or to national level data as available.

For analysis of growth over time, we are able to show development on outcomes on which we have three data points collected over time. These curves describe the rate of children's growth on the outcome of interest and how that changes over time (that is, the acceleration/deceleration of change rate). Because children/families are nested within programs, children/families in the same program are likely to be more similar to each other than to children/families in other programs. Using multilevel modeling can improve the precision of estimates of the associations of change/growth with child/family and program characteristics by taking the nested data structure into account. Therefore, we estimated a three-level model for each child/family outcome: within individual over time (level 1), between individuals within programs (level 2), and between programs (level 3). We use Hierarchical Linear modeling (HLM) software (Raudenbush and Bryk 2002) for these estimates (Appendix D provides further details of this modeling approach).

The level-1 model estimates change over time within individuals. It demonstrates the change trajectory over time for each of the outcomes—whether it is linear or nonlinear. The level-2 model estimates variation between individuals within programs. By including the child/family characteristics in the model, we examine whether the change trajectories of the outcomes differ by these variables. The level-3 model estimates variation between programs. The purpose of the program-level model is to examine how program characteristics are related to individual change in the outcome measures.

The main limitation of the growth curves is that changes over time in the rate of growth cannot be causally attributed to the program and may in fact be due to changes in the composition of the sample. For example, if children with challenging behavior are more likely to leave the program than

other children, then rates of problem behavior will appear to decrease rapidly when actual behavior may not have changed at all. We attempt to address this by conducting sensitivity tests using children with data at all three time points to compare the shape of those curves to those including children who are missing some data points. In addition, we include important child (e.g., age, cohort, health, baseline ASQ) and family characteristics (e.g., baseline maternal risk, parenting stress), in the models.

Predictors of Home Visit and Classroom Quality

We used three-level HLM models to describe predictors of quality of home visits (Visitor Strategies and Visitor Effectiveness) and classrooms (Engaged Support for Learning and Emotional and Behavioral Support) using child, family, staff, classroom or home visit, and program characteristics. In each model, level-1 represents how quality varies over time within individual staff members, level-2 describes how quality varies between different staff members in the same program, and level-3 describes how quality varies between staff members in different programs. Our models account for the fact that multiple assessments of the same staff member over time are not independent observations. We also adjust for the shared experiences of staff members in the same program whose observed quality may be more similar than staff from different programs.

Predictors of Early Head Start Participation

We used multilevel models to examine the relationships among program, staff, child and family characteristics and Early Head Start participation (defined as length of program enrollment, staff ratings of family involvement, and home visits and center days received). For length of program enrollment, we fit two-level HLM models to examine how enrollment varies both within and between Early Head Start programs. In each model, level-1 represents how length of enrollment varies among families in the same program and whether differences are due to child or family characteristics. Level-2 describes variation between programs and allows exploration of how program characteristics are related to participation. This analytic approach enables us to account for the fact that our study sample includes children and families who attend the same programs and may have more similar participation patterns compared to families in different programs. For measures of participation assessed at multiple time points, we fit 3-level models to examine how participation varies over time within individual children/families in addition to examining within- and between-program variation. In each model, level-1 represents how participation of an individual child or family varies over time. Level-2 describes how participation differs among families in the same program and whether differences are due to child or family characteristics. Level-3 describes variation between programs and allows exploration of how program characteristics are related to participation.

Relationships Between Participation and Outcomes

In estimating the relationship between Early Head Start participation and outcomes, it is important to account for the fact that differences we observe between families may be due to differences in their participation or due to differences in other characteristics that resulted in differences in participation. The uniquely rich data collected for Baby FACES enable us to control for many of the factors influencing participation and outcomes so that we can more clearly assess our relationships of interest.

Our approach has several steps. As a first step in analysis, we assess the bivariate relationship between each participation and outcome measure. When we find significant bivariate associations, we proceed with multivariate analyses that include a comprehensive set of child, family, and program characteristics as covariates. We conduct sensitivity analyses to gauge the extent of selection bias in

our estimates and as a check that our results are not sensitive to functional form assumptions. Detailed information on our estimation methods is included in Chapter IX and in Appendix D.

III. HOW EARLY HEAD START CHILDREN AND FAMILIES ARE FARING IN THEIR LAST YEAR IN THE PROGRAM

In the first two Baby FACES reports, we described how 1-year-olds enrolled in the Early Head Start program in spring 2009 were functioning at ages 1 and 2. In this final report, we describe the experiences of the 1-year-old Cohort and Newborn Cohort children who continue to be enrolled in Early Head Start in spring 2011 and 2012 when they are 3 years old. We draw on multiple data sources to capture children's development in different contexts, including reports from parents, Early Head Start staff, and assessors; direct child assessments; and video-recorded adult-child interactions during the in-home assessment. We also assess parenting behaviors and home and neighborhood environment through parent interviews, parent self-administered questionnaires (SAQ), assessor observations, and coded semi-structured parent-child play interactions.

This chapter reports on how the children and families are faring in their last year in Early Head Start. The first section of the chapter describes children's physical health and access to health care using information obtained through parent interviews and direct child assessments. The second section focuses on children's general development as measured by the Ages & Stages Questionnaires, Third Edition (ASQ-3), which parents completed as part of the SAQ during administration of the child assessments. We then describe children's language development at age 3, based on reports from parents and Early Head Start staff, direct child assessments, and video-recorded interactions with the child, followed by children's social-emotional development as rated by parents, Early Head Start staff, and assessors. The final section of the chapter is devoted to parenting and the home and neighborhood environment. To place findings in context, at times we make comparisons to other research with similar populations or using similar measures (such as the EHSREP, FACES, national norms, and so on). Any other comparisons within the Baby FACES sample, such as across years of the study, or between groups, are mentioned if they are significant at the .05 level or less.

Parents Report that Most 3-Year-Olds in Early Head Start Are Healthy and Have Access to Health Care

In the spring 2011 and 2012 interviews, we asked parents of 3-year-olds to report on the children's general health status, health care access, and disability evaluation and services since the last interview. We also collected information on child height and weight during one-on-one assessments completed with each child.

Most 3-Year-Olds Are in Good Health, but One-Third Are Obese or at Risk for Obesity

The majority of 3-year-olds enrolled in Early Head Start are healthy. Table III.1 presents the weighted means and percentages of the measures of child health. According to parents' reports, 80 percent of 3-year-olds have excellent or very good health; only 4 percent have fair or poor general health. As reported on the five-point rating scale ranging from poor (1) to excellent (5), the mean parent rating is 4.3, suggesting that, overall, children's general health status is between very good and excellent.

We calculated children's body mass index (BMI) for age percentile based on the 2000 CDC growth charts (Kuczmarski et al. 2002) using height and weight data collected during the direct child assessments. According to expert committee recommendations (Barlow et al. 1998), children with a BMI for age at the 95th percentile or greater are identified as obese; those between the 85th and 94th percentiles are overweight; and those with BMIs lower than the 5th percentile are underweight. Among 3-year-olds who were enrolled in Early Head Start in spring 2011 and 2012, the prevalence of

obesity is 16 percent—which is similar to the 16 percent prevalence of obesity among a nationally representative sample of 3-year-old children found in FACES (Aikens et. al. 2010) and 18 percent rate of 4-year-old children found in the ECLS-B (Anderson and Whitaker 2010); about 16 percent of children are identified as at risk for obesity; 6 percent are underweight.

Table III.1. The Majority of 3-Year-Olds are Healthy

Child Characteristics	Weighted Means or Percentages (Standard Error)
Health Status	4.3 (0.06)
Excellent or very good	79.6 (2.71)
Fair or poor	3.5 (0.99)
Child BMI	
Child is obese (BMI ≥ 95th percentile)	16.1 (1.93)
Child is overweight (BMI ≥ 85th and < 95th percentile)	16.3 (1.88)
Child is underweight (BMI < 5th percentile)	5.8 (1.18)
Sample Size	
Parent interview	**410-438**

Source: Spring 2011 and 2012 Parent Interview.

Note: Sample includes both the 1-year-old Cohort and the Newborn Cohort at age 3.

BMI =body mass index.

Most 3-Year-Olds Have Access to Health Care

Access to health care and health insurance coverage is important for promoting good health in children. Table III.2 summarizes information obtained through parent interviews regarding what health care services children have received and their health insurance coverage.

At age 3, essentially all (99 percent) Early Head Start children received some type of health services, including doctor or dentist visits, and immunizations. Only 2 percent of Early Head Start children do not have a regular health care provider—this is lower than the 6 percent of children nationally who do not have a regular source of health care (Federal Interagency Forum on Child and Family Statistics 2010). Approximately 15 percent of parents reported that Early Head Start helped them find regular health care. All 3-year-olds have visited a doctor since the last interview; three-quarters (75 percent) of children have visited a dentist. Ninety-one percent of the children have had a well-child checkup at least once since the last interview; about 96 percent had their last checkup fewer than 12 months ago. Approximately 94 percent of 3-year-olds are reported as being "completely up to date" with immunizations.

According to parents' reports, approximately 6 percent of children have been hospitalized since the last interview, and 36 percent of children have visited the emergency room (ER). A high fever is the most commonly reported reason for ER visits (47 percent of those visiting the ER). The next most common reasons noted are breathing problems (38 percent) and an accident or injury (38 percent). About one-third (35 percent) of children visited the ER due to ear infection. Approximately 18 percent sought emergency treatment for dehydration or diarrhea. Two percent of parents reported jaundice as the reason for their ER visit.

Children with health insurance coverage are more likely to have a regular and accessible source of health care. Among 3-year-olds who were enrolled in Early Head Start in spring 2011 and 2012, only 2 percent do not have any health insurance coverage. This is considerably lower than the national rate of 10 percent (Federal Interagency Forum on Child and Family Statistics 2010). Most children (89

percent) are covered by public plans; more than one-third (37 percent) are covered by private health insurance.[8] Approximately three-quarters (76 percent) of children have dental insurance coverage. Nine and 15 percent of parents reported that Early Head Start helped them find health and dental insurance, respectively.

Table III.2. The Majority of 3-Year-Olds Have Access to Health Care

Child Characteristics	Weighted Percentages (Standard Error)
Has a Regular Health Care Provider	98.0 (0.71)
Early Head Start Helped Parent Find Regular Health Care	15.4 (2.14)
Received Any Health Services	99.4 (0.39)
Child Visited	
A doctor for a checkup	100.0 (0.00)
A dentist	74.8 (3.39)
Child's Last Regular Doctor Checkup Was Fewer than 12 Months Ago	95.8 (1.17)
Frequency of Well-Child Checkups Since Last Interview	
Never	8.9 (2.40)
Once or twice	65.2 (3.20)
3–4 times	17.6 (2.08)
5–9 times	4.2 (1.00)
10 times or more	4.1 (1.04)
Child's Immunization Status Is "Completely Up to Date"	94.0 (1.65)
Hospitalized Since Last Interview	6.0 (0.97)
Has Had an Emergency Room (ER) Visit	35.8 (2.79)
Reason for ER Visit [a]	
High fever	47.4 (5.15)
Breathing problems	37.7 (4.36)
Accident or injury	37.6 (4.58)
Ear infection	35.2 (4.25)
Dehydration or diarrhea	17.9 (3.90)
Jaundice	2.0 (1.05)
Child's Health Insurance Status[b]	
Private health insurance plan	36.8 (3.01)
Public/government insurance	88.6 (2.07)
No health insurance	1.7 (0.67)
Child has dental insurance coverage	75.8 (2.81)
Early Head Start Helped Parent Find Health Insurance[c]	9.2 (1.88)
Early Head Start Helped Parent Find Dental Insurance[d]	14.5 (2.44)
Sample Size	
Parent interview	**400-410**

Source: Spring 2011 and 2012 Parent Interview.

Note: Sample includes both the 1-year-old Cohort and the Newborn Cohort at age 3.

[a]Only for those who have had ER visits (N = 153).
[b]The estimates are not mutually exclusive and hence sum to more than the estimated percentage of children with coverage.
[c]Only for those with health insurance (N = 397).
[d]Only for those with dental insurance (N = 312).

[8] Children could be covered by both types of plans; therefore, the estimates sum to more than the estimated percentage of children with coverage.

According to Parents' Reports, more than Half of Children with Special Needs Received Disability Services

Parents of 3-year-olds also reported on children's special needs and early intervention services since the last interview (Table III.3). More than one-quarter (27 percent) of parents reported their children had special needs;[9] among these children, 56 percent had been evaluated for disabilities since the last interview. The most common special needs reported by parents are hearing and speech problems (14 percent); 72 percent of these children had been evaluated for such problems (Figure III.1). Behavioral/attention problems were reported for 10 percent of children, with one third (33 percent) of these children evaluated for these problems. Sleep apnea was reported for 5 percent of children; 58 percent of these children were evaluated. Two to 4 percent of children were reported as having motor problems and developmental delay; about three-quarters (77 and 76 percent for developmental delay and motor problems, respectively) of these children had been evaluated. Two percent of children experienced vision problems or below-normal activity level, with 68 percent of children who had vision problems and 12 percent of children who had below-normal activity level evaluated for such problems. Approximately one-third (34 percent) of all children had been evaluated for disabilities since birth.

Overall, 59 percent of children who were evaluated for special needs since the last interview have received disability services; 80 percent of these parents reported that Early Head Start helped them obtain these services. Eighty-seven percent of children who have been evaluated for special needs since birth have received disability services, and all of these parents reported that Early Head Start helped them obtain these services. Less than half of children with special needs (41 percent) are reported to be currently participating in an early intervention program. About one-third (33 percent) of all parents reported that children have an Individualized Education Plan (IEP) or Individualized Family Service Plan (IFSP).

[9] This finding is based on parent reports, regardless of whether the child has a formal diagnosis. Specifically, we asked parents if their child had any of the disabilities, and only those who answered yes were asked if the child was evaluated.

Table III.3. Receipt of Early Intervention Services at Age 3

Characteristics	Weighted Percentages (Standard Error)
Child Has Any Disabilities According to Parent Report	26.7 (2.37)
Hearing or speech problems	13.9 (1.93)
Behavioral or attention problems	10.3 (1.89)
Sleep apnea	4.8 (1.31)
Developmental delay	3.7 (0.96)
Motor problems	2.4 (0.92)
Below-normal activity level	1.6 (0.68)
Vision problems	2.1 (0.77)
Child Has Been Evaluated for any Disabilities[a] Since Last Interview	55.9 (5.15)
Hearing or speech problems	72.2 (5.93)
Developmental delay	76.8 (11.35)
Motor problems	75.8 (2.35)
Sleep apnea	57.8 (9.21)
Behavioral or attention problems	33.2 (7.90)
Below-normal activity level	11.8 (6.54)
Vision problems	67.9 (10.31)
Child Has Ever Been Evaluated for any Disabilities	34.2 (3.16)
Child Has Received Disability Services[b] Since Last Interview	59.2 (7.45)
Early Head Start Has Helped Family and Child Obtain Disability Services[c] Since Last Interview	80.3 (8.14)
Child Has Ever Received Disability Services[d]	87.4 (5.18)
Early Head Start Has Ever Helped Family and Child Obtain Disability Services[e]	100.0 (0.00)
Child Currently Participating in an Early Intervention Program[b]	40.9 (4.95)
Child has an IEP/IFSP	33.1 (2.70)
Sample Size	
Parent interview	**402-410**
Reported by parent as having disabilities	**9-106**
Received disability services since last interview	**63**

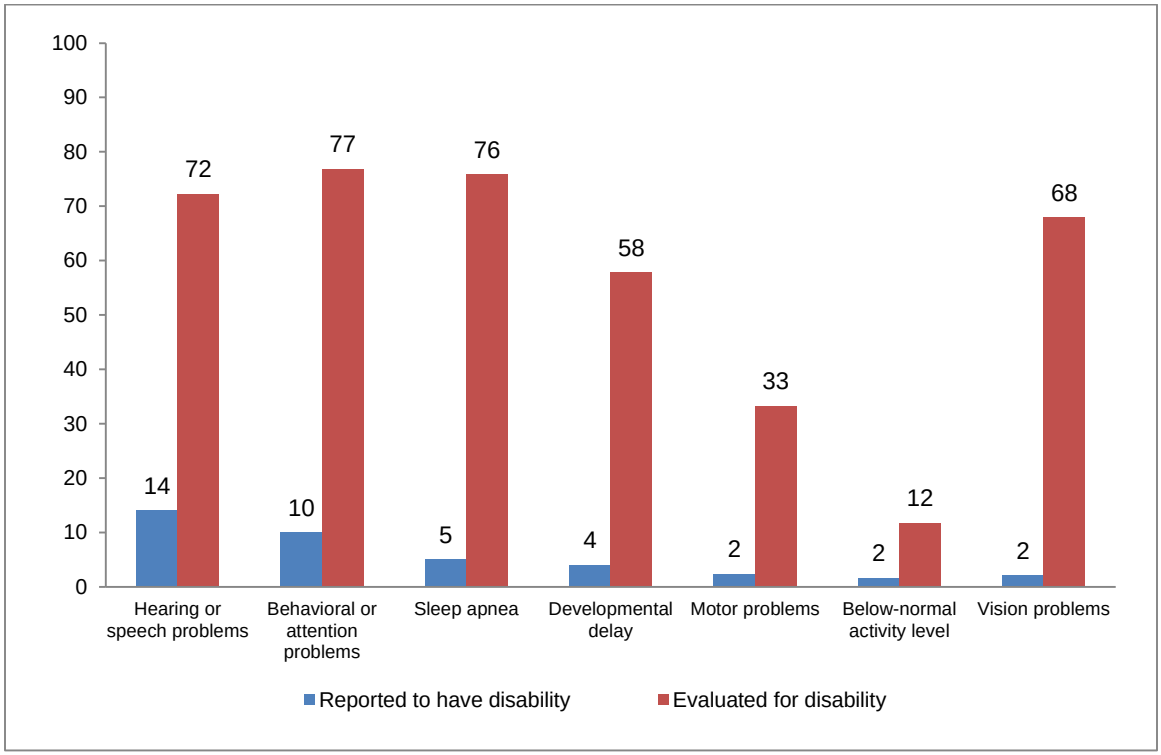
Sources: Spring 2009 to 2012 Parent Interview.

Note: Sample includes both the 1-year-old Cohort and the Newborn Cohort at age 3.

[a]Among those who were reported by parents as having disabilities (N = 106).
[b]Among those who were reported by parents as evaluated for disabilities since last interview (N = 63).
[c]Among those who have received disability services since last interview (N = 42).
[d]Among those who were reported by parents as evaluated for disabilities since birth (N = 73).
[e]Among those who have received disability services since birth (N =55).

IEP = individualized education program or plan; IFSP = individual family service plan.

Figure III.1. Percent of Children Reported to Have a Disability or Have Been Evaluated for a Disability



On Average, 3-Year-Old Early Head Start Children Scored Similarly to the National Norms in General Development

Parents of 3-year-olds completed the ASQ-3 (Squires et al. 2009), which measures children's cognitive, communication, social, and motor development, and identifies children who may be at risk in these developmental areas (Table III.4). Less than one-quarter (23 percent) of parents completed the ASQ-3 in Spanish. As defined in the ASQ-3 user's manual, children who score two standard deviations or more below the mean might be at risk and require further assessment. Children who score in the monitoring zone (between one and two standard deviations below the mean) need further monitoring and may benefit from practicing skills in a specific area of development. (See Box III.1 for a description of the ASQ-3 scores and scoring procedures.)

Table III.4. Children's General Developmental Status at Age 3: ASQ-3

Measures	Weighted Means or Percentages (Standard Error)
ASQ-3 Age-Specific Raw Score[a]	
Communication	51.0 (0.61)
Gross Motor	53.8 (0.44)
Fine Motor	44.7 (0.77)
Problem Solving	51.3 (0.46)
Personal-Social	53.9 (0.41)
Total Score	254.8 (1.98)
ASQ Cut-Off Score (2SD below the mean or lower)	
Communication	0.6 (0.41)
Gross Motor	5.5 (1.23)
Fine Motor	5.9 (1.54)
Problem Solving	5.6 (1.25)
Personal-Social	4.2 (1.42)
ASQ in the monitoring zone (1-2SDs below the mean)	
Communication	9.4 (1.44)
Gross Motor	11.2 (1.32)
Fine Motor	13.8 (1.83)
Problem Solving	12.9 (1.82)
Personal-Social	6.1 (1.17)
ASQ-3 Raw Scale Score[b]	
Communication	85.0 (0.95)
Gross Motor	80.2 (0.64)
Fine Motor	64.7 (1.03)
Problem Solving	83.6 (0.84)
Personal-Social	89.1 (0.64)
Total Scale score	402.9 (3.19)
ASQ-3 IRT Score	
Communication	79.2 (0.90)
Gross Motor	82.7 (0.82)
Fine Motor	73.9 (0.77)
Problem Solving	79.5 (0.84)
Personal-Social	83.4 (0.66)
Sample Size	
Parent SAQ	**372-450**

Source: Spring 2011 and 2012 Parent Self-Administered Questionnaire (SAQ).

Note: Sample includes both the 1-year-old Cohort and the Newborn Cohort at age 3.

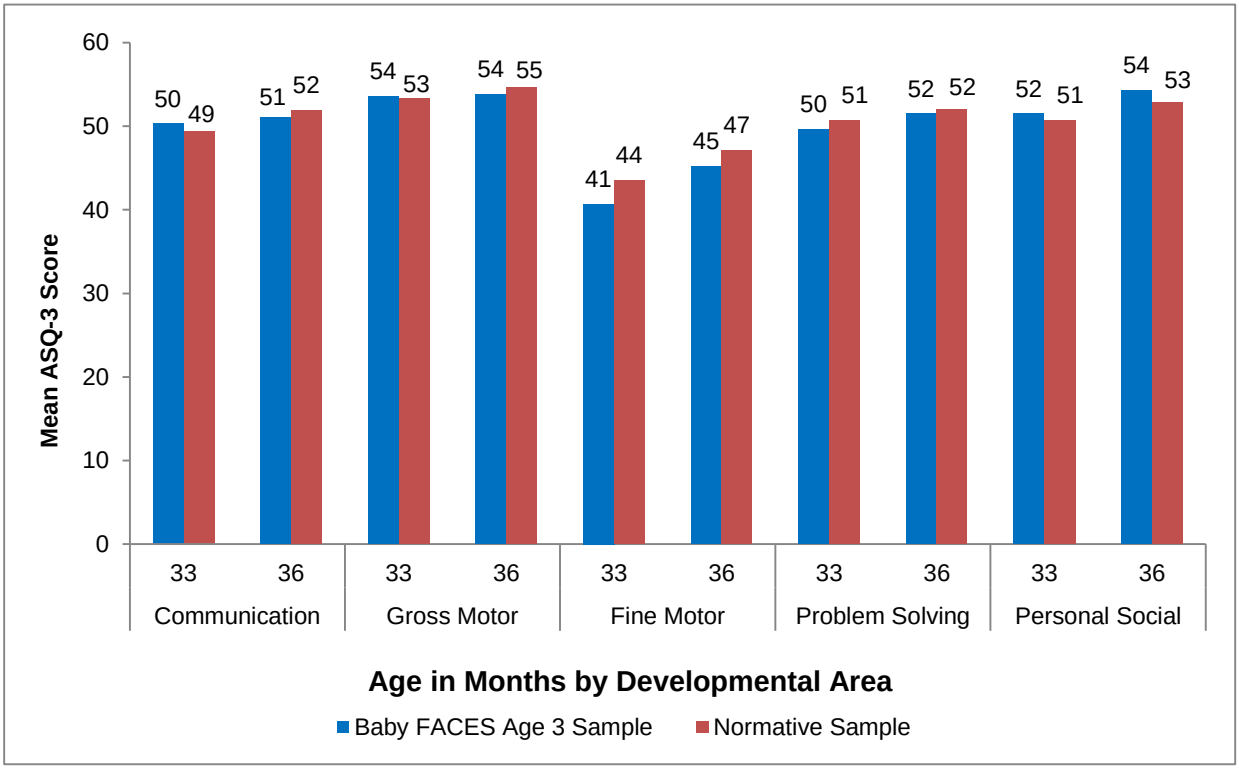
[a]Includes only items on the age-specific forms.
[b]Includes all items across age forms.

ASQ-3 = Ages & Stages Questionnaires (Third Edition).

On average, 3-year-olds enrolled in Early Head Start in spring 2011 and 2012 are catching up with same-age children nationally in each of the developmental areas. The average scores on the age-specific forms range from 51–54 (out of 60) in Communication, Gross Motor, Problem Solving, and Personal-Social, with the lowest mean score in Fine Motor (about 45 out of 60). The average ASQ-3 total score for 3-year-olds is 255 out of 300; which is the same as the average total score in the Baby FACES sample. At age 3, Early Head Start children scored similarly to the normative sample in all five areas of development: Communication, Gross Motor, Fine Motor, Problem Solving and Personal-Social (Figure III.2).[10]

[10] Differences are not based on tests of significance.

Figure III.2. ASQ Scores at Age 3 Relative to Norms



Approximately 4 to 6 percent of 3-year-olds in Early Head Start scored below the at-risk cutoff in Gross Motor, Fine Motor, Problem Solving, and Personal-Social. Less than 1 percent (0.6 percent) of children scored below the cutoff in Communication. Children were most likely to score in the monitoring zone in Fine Motor and Problem Solving (14 and 13 percent, respectively); 11 percent scored in the monitoring zone in Gross Motor; 9 percent in Communication, and 6 percent in Personal Social.

Box III.1. Measure of Child General Development

Several factors were considered in the selection of measures, including the psychometric strength of the measure, availability in Spanish, the likelihood that they could be adopted by Early Head Start programs or were already in wide use among programs, comparability to other large scale studies and feasibility for use in this large scale study. The measure below is considered a screener which is meant to identify children at an elevated risk for delayed development and initiate further, in-depth assessments. In addition to providing information on children whose screenings indicate problems, assessments also provide information about children's abilities and skills that programs can use to design interventions that support their development.

Ages & Stages Questionnaires, Third Edition (ASQ-3; Squires et al. 2009) is a parent-report tool for screening infants and young children for developmental delays. The ASQ-3 includes 21 questionnaires that are appropriate for children aged 1 month through 5-1/2 years. These surveys focus on the assessment of five key developmental areas: (1) Communication, (2) Gross Motor, (3) Fine Motor, (4) Personal-Social, and (5) Problem Solving. Parents are asked to respond "not yet," "sometimes," or "most of the time" to questions such as, "Does your child make sentences that are three or four words long?" There are six items in each of the five developmental areas. The raw score in each developmental area can range from 0 to 60, and the ASQ-3 total area score can range from 0 to 300. For 3-year-olds in the third round of Baby FACES data collection, the unweighted mean for the ASQ-3 total area score is 253.2 (SD = 39.0). For the developmental area scores, the unweighted means range from 44.7 to 53.8 (with SDs ranging from 7.9 to 14.7).

- As a screening tool, the age-specific raw score for each of the ASQ-3 developmental areas has a ceiling problem, because many children score at the maximum. This tendency limits our ability to examine change over time in each of the child's developmental areas. To address this problem, we developed in the third round of Baby FACES data collection a form that included all of the items required by the two forms we need to cover the range of ages of the 1-year-old Cohort children at age 3 (33 and 36 months) as well as items from the 42-month form. This approach is different from how the ASQ is typically used, but it enabled us to look at child functioning in a more dimensional or continuous way rather than just whether they had a developmental concern or not. We developed this approach in collaboration with the ASQ test developer to ensure the resulting adaptation of the instrument was acceptable. We analyzed items across the forms and create both the age-specific area scores as well as summary (scale and IRT) scores we could then analyze, controlling for child age. The scale scores can range from 0 to 100 for Communication, Problem Solving, and Personal-Social; 0 to 90 for Gross Motor and Fine Motor. The unweighted means for the ASQ-3 scale scores for the 3-year-olds in spring 2011 are 84.9 (SD = 16.6) for Communication, 64.3 (SD = 20.0) for Fine Motor, 88.7 (SD = 12.9) for Personal-Social, 79.5 (SD = 12.1) for Gross Motor, and 83.3 (SD = 17.0) for Problem Solving. The unweighted mean for the total scale score across domains is 400.3 (SD = 62.5) out of 480.

- Using all the items across the 33- to 42-month forms from the third round of data collection as well as all the items across the 22- to 30-month forms from the second round of data collection, we create the IRT scores that take item difficulties into account. The IRT scores range from 0 to 100. The unweighted means of the IRT scores are 79.5 (SD = 14.1) for Communication, 73.4 (SD = 14.0) for Fine Motor, 81.8 (SD = 13.3) for Gross Motor, 83.0 (SD = 12.8) for Personal-Social, and 79.1 (SD = 13.6) for Problem Solving at age 3.

- The developer-derived cutoff points, which vary by age and indicate a need for further assessment, are two standard deviations below the mean in each developmental area. Children scoring two standard deviations below the mean or lower are in the at-risk range. For example, the cutoff in Communication is 25.36 for the 33-month form and 30.99 for the 36-month form. The cutoff of two standard deviations has a sensitivity and specificity of 0.86. In other words, those whose scores are two standard deviations below the mean or lower have an 86 percent chance of being identified for further assessment. Children whose scores fall in the monitoring zone—defined by the ASQ-3 authors as one to two standard deviations below the mean—might benefit from practicing skills in a specific area of development. As would be expected, the cutoff point of one standard deviation has a high sensitivity (0.98) but a low specificity (0.59; Squires et al. 2009). (Appendix C details the ASQ-3 norming sample and the psychometric properties observed in this study.)

Most Early Head Start Children Demonstrate Good Language Skills at Age 3

As in the spring 2010 data collection for 2-year-olds, we draw on data from multiple sources to document 3-year-old children's language development across different contexts (Table III.5). One of the language measures used was the MacArthur-Bates Communicative Development Inventories-III (CDI-III) for children 30–37 months of age (Fenson et al. 2007). Early Head Start staff who spoke Spanish also completed the Spanish CDI-III for Spanish-speaking children (Vagh, Mançilla-Martinez, and Pan, unpublished manuscript; see Box III.2 for a description of the CDI scores and scoring procedures).

We also administered the Preschool Language Scale-Fourth Edition (PLS-4; Zimmerman et al. 2002) and the Peabody Picture Vocabulary Test-Fourth Edition (PPVT-4; Dunn and Dunn 2007) to children during one-on-one direct child assessments conducted as part of the in-person home visit to assess auditory comprehension and receptive vocabulary, respectively. We also assessed children's expressive communication while engaged in a 6-minute play-based communication task with the assessor (Early Communication Indicator [ECI]; Carta et al. 2010; Luze et al. 2001). Box III.2 provides descriptions and scoring procedures for the PLS-4, PPVT-4, and ECI.

Across most measures and reporters (parents, staff, and direct child assessments), children perform close to (but often slightly below) same age children. In the next sections we describe findings for each of the measures that we used with 3-year-olds.

Box III.2. Measures of Child Language Development

MacArthur-Bates Communicative Development Inventories (CDI) (Fenson et al. 2000) is designed to assess children's early receptive and expressive language and communication skills through parent report. At round 3 of Baby FACES data collection, Early Head Start staff (teachers and home visitors) completed the CDI-III vocabulary checklist for the 3-year-olds. The 100-item CDI-III vocabulary checklist is a short measure of expressive vocabulary for children 30–37 months of age. We also asked Early Head Start staff to report on 3-year-olds' receptive vocabulary using the checklist. Teachers and home visitors who reported they spoke Spanish completed the English CDI-III and the Spanish CDI-III (Vagh, Mançilla-Martinez, and Pan, unpublished manuscript) for children identified as understanding Spanish. Two measures were derived from each of the forms:

- Vocabulary Comprehension measures the number of words the child understands. Teachers and home visitors were asked whether the child understands or, both understands and says each of 100 specific words.
- Vocabulary Production measures the number of words in the child's spoken vocabulary. Early Head Start teachers and home visitors reported whether the child understands and says each of 100 specific words.

In addition to staff reports, parents also report on children's English or Spanish Vocabulary Production using the CDI-III in a self-administered questionnaire. The raw scores for both Vocabulary Comprehension and Vocabulary Production range from 0 to 100. Using the staff reports, the unweighted means are 64.9 (SD = 25.2) and 71.1 (SD = 26.2) for English and Spanish Vocabulary Comprehension, respectively, and 41.4 (SD = 27.2) and 41.2 (SD = 29.4) for English and Spanish Vocabulary Production, respectively. Using the parent reports, the unweighted means are 58.3 (SD = 28.9) and 55.5 (SD = 27.3) for English and Spanish Vocabulary Production, respectively. Appendix C details the CDI norming sample and the psychometric properties observed in this study.

Preschool Language Scale—Fourth Edition (PLS-4; Zimmerman et al. 2002a, 2002b) is a direct child assessment used to evaluate receptive and expressive language skills, as well as understanding and use

of grammatical rules for children from birth to 6 years of age. It is composed of two subscales: Auditory Comprehension (AC) and Expressive Communication (EC). We used the AC subscale for both of the English and Spanish editions of PLS-4 at round 3 of Baby FACES data collection. The tasks in the AC subscale are designed to assess skills that are important for language development (for example, following directions with cues and appropriate object play). The unweighted means of standard scores are 97.4 (SD = 15.4) for English and 95.7 (SD = 18.0) for Spanish at age 3.

- In discussion with the test publisher, we developed a procedure to derive a "conceptual score" for DLLs, giving children credit for their knowledge of both English and Spanish. For Spanish-speaking children, we also administered in English the items that they did not pass in Spanish until the children reached an English ceiling. Thus, we are able to calculate the conceptual scores by giving children credits for items that they answered correctly in Spanish and/or English and derive the bilingual standard scores using the norms for the Spanish Edition. The unweighted mean is 103.1 (SD = 16.4) for the bilingual standard scores at age 3.

Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4; Dunn & Dunn, 2007) is a norm-referenced standardized test designed as a measure of receptive vocabulary and screening test for verbal ability and is suitable for a wide range of ages, from 2½ through adulthood. We administer the PPVT-4 to all children regardless of their primary language at round 3 of Baby FACES data collection. Children are asked 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 unweighted mean of the PPVT-4 standard scores is 91.4 (SD = 13.8) at age 3.

Early Communication Indicator (ECI; Luze et al. 2001; Carta et al. 2010) is a semi-structured, play-based assessment designed to measure the expressive communication of infants and toddlers between the ages of 6 and 36 months along four key skill elements: (1) gestures, (2) vocalizations, (3) single-word utterances, and (4) multiple-word utterances. (Appendix C defines each of the skill elements.) As part of the assessment activities conducted with 3-year-olds, assessors administered the ECI, which was video-recorded for later coding by staff at Mathematica. Coders record the frequency of occurrence of each skill element over the six-minute assessment. Observed instances are combined to yield a total communication score that reflects the weighted combination of the child's gestures, vocalizations, and single- and multiple-word utterances. (The latter two are given weights of two and three, respectively, to account for the greater complexity of skill associated with their use.) Total weighted scores are converted to a rate score that reflects the number of communicative bids per minute over the course of the six-minute play assessment (unweighted mean = 16.6; SD = 8.6. An age-based, standardized score with a mean of 100 (SD = 15) is also computed. The unweighted mean of standardized scores is 93.0 (SD = 15.8). Two cutoff scores identify children with (or at risk for) expressive language delays. Children scoring between one and one and a half standard deviations below the mean are in the at-risk range; those with scores one and a half standard deviations below the mean or lower are identified as demonstrating delays in expressive language (Greenwood et al. 2006, 2010).

Table III.5. Children's Language Development at Age 3

Measures	Weighted Mean or Percentage (Standard Error)
Staff-Reported CDI	
English CDI raw score	
Vocabulary comprehension	64.7 (2.06)
Vocabulary production	42.6 (1.77)
Spanish CDI raw score	
Vocabulary comprehension	72.3 (2.69)
Vocabulary production	41.5 (4.76)
Parent-Reported CDI Raw Score	
English vocabulary production	56.9 (1.84)
Spanish vocabulary production	57.8 (2.13)
Combining words	
Overall	
Not yet	4.8 (1.55)
Sometimes	18.5 (1.97)
Often	76.8 (2.62)
English	
Not yet	5.6 (1.84)
Sometimes	19.5 (2.21)
Often	74.9 (3.02)
Spanish	
Not yet	0.7 (0.70)
Sometimes	13.1 (3.46)
Often	86.2 (3.61)
Child's Ability to Speak Language Used Most Often[a]	
Speaks only a few words or phrases	4.8 (1.06)
Speaks it but has limited vocabulary	18.8 (2.13)
Speaks it and has good vocabulary	76.4 (2.25)
Child's Ability to Understand Language Used Most Often[b]	
Understands only a few words	2.2 (0.69)
Understands general idea of what was said	8.0 (1.65)
Understands most or all of what was said	89.8 (1.75)
Child's Ability to Understand Language Used Most Often	
Better than ability to speak language	19.3 (2.05)
About same as ability to speak language	74.6 (2.21)
Worse than ability to speak language	6.0 (1.37)
PLS-4 Standard Score	
English	97.0 (1.17)
Spanish	97.8 (1.63)
Bilingual	105.2 (1.60)
PPVT-4 English Standard Score	
Overall	91.3 (0.85)
DLL	86.3 (1.57)
Non-DLL	93.4 (0.95)
ECI	
Standard score	93.3 (0.84)
ECI language delay (percentage 1.5 SDs below the mean or lower)	17.6 (1.82)
ECI at-risk for language delay (percentage 1 to 1.5 SDs below the mean)	15.8 (1.59)
Sample Size	
Parent SAQ English CDI	**344**
Parent SAQ Spanish CDI	**102**
SCR CDI English	**484**
SCR CDI Spanish	**92**
ECI Assessment	**450**
Parent interview	**417-419**

Sources: Spring 2011 and 2012 Staff-Child Report (SCR), Parent Self-Administered Questionnaire (SAQ), Direct Child Assessment, ECI Assessment, and Parent Interview.

Note: Sample includes both the 1-year-old Cohort and the Newborn Cohort at age 3.

[a]If the parent reported the child spoke more than one language equally as often, the child's highest reported speaking ability of all languages was used.

[b] If the parent reported the child spoke more than one language equally as often, the child's highest reported comprehension ability of all languages was used.

CDI = MacArthur-Bates Communicative Development Inventories; PLS-4 = Preschool Language Scale (Fourth Edition); PPVT-4 = Peabody Picture Vocabulary Test (Fourth Edition); DLL = Dual language learner; ECI = Early Communication Indicator.

Parents Reported Better Expressive Language Development in Children than Did Early Head Start Staff

According to Early Head Start staff reports, 3-year-old children understand 65 of the 100 English words in the CDI-III vocabulary checklist and say 43 English words, on average (Table III.5). Children from homes in which English is the only language spoken scored higher on the English CDI than did DLLs—70 versus 54 on vocabulary comprehension and 49 versus 31 on vocabulary production. Spanish-speaking children (with a Spanish-speaking staff member) understand 72 Spanish words and say 42 Spanish words.

According to parent reports, English-speaking children say 57 of the 100 English words included in the CDI-III vocabulary checklist, and Spanish-speaking children say 58 Spanish words, on average.[11] These scores are higher than those reported by Early Head Start staff. Parents also reported on whether children can combine words. Overall, approximately three-quarters (77 percent) of parents reported that their child can often combine words, while 5 percent of the children have not yet begun to combine words, and 19 percent can sometimes combine words.[12] When examining the percentages by children's language, Spanish-speaking children are more likely than their English-speaking peers to have begun to combine words (99 percent versus 94 percent) and to be able to combine words often (86 percent versus 75 percent). About one-fifth of English-speaking children (20 percent) can sometimes combine words and about one-tenth of Spanish-speaking children can (13 percent).

Parents Reported that most Children Demonstrate Good Language Comprehension and Production Skills

Parents reported that most 3-year-olds have strong speaking ability in the language they speak most often. Approximately three-quarters (76 percent) of parents reported that their child has a good vocabulary. One-fifth (19 percent) said their child has a limited vocabulary. Only five percent of parents report that their child can speak only a few words or phrases in the language they speak most often.

Parents also reported that most children can understand well the language they speak most often. The majority of parents (90 percent) said their child can understand most or all of what is being said to them; 8 percent reported their child can understand the general idea of what is being said to them. Two percent of parents reported that their child can understand only a few words—most (72 percent) of these children were reported as having special needs.

In addition, most parents reported that their child understands the language at a level similar to, or better than, his or her ability to speak the language. Three-quarters of parents rated their child's speaking and comprehension abilities at the same level. An additional 19 percent rated their child's comprehension as better than his or her speaking ability.[13]

[11] Parents did not report on vocabulary comprehension.

[12] We did not ask Early Head Start staff to report on combining words.

[13] About 6 percent of parents said that their child speaks his or her first language at a level better than he or she understands it (which may indicate that parents did not understand the question).

Early Head Start Children's Auditory Comprehension Approaches National Norms, While Receptive Vocabulary and Expressive Language Skills Lag Behind Peers'

Direct child assessments of language ability include the PLS-4, PPVT-4, and the ECI, a 6-minute play-based assessment completed by the assessor. This play task was video-recorded for later coding by a trained Mathematica coding staff member.

Three-year-old Early Head Start children score close to national norms on the PLS-4 Auditory Comprehension subscale. The mean scores on the English and Spanish PLS-4 Auditory Comprehension (means = 97 and 98, respectively) are about one-fourth of a standard deviation below the national norms.[14] Spanish-speaking children's conceptual scores on the PLS-4 are about one-fourth of a standard deviation above national norms (mean = 105).

Children's performance on the PPVT-4 lags behind the normative sample. Overall, the mean score (mean = 91) is one-half of a standard deviation below the national mean, with children from homes where English is the only language spoken scoring higher than DLLs (mean = 93 versus 86).

Similarly, the expressive communication skill of 3-year-old children on the ECI is approximately one-half of a standard deviation below the standardized mean for children at this same age (mean = 93). When examined by child age in months at the time of the assessment (34 to 38 months), standardized mean scores range from 90 to 94.[15] Overall, the total weighted rate per minute communication score averaged 16.6 (SD = 8.6). Although there was considerable variability in children's scores (ranging from 0 to 45 communicative bids per minute), mean scores are lower than those reported in normative samples of children at this same age (20.2 communicative bids per minute; Greenwood et al. 2006). According to established benchmarks for children at this age (Carta et al. 2010; Greenwood et al. 2006), 33 percent of 3-year-olds in Baby FACES demonstrate or are at risk for delays in expressive language skill. Specifically, 18 percent of children score at or below the threshold for delays in expressive language (1.5 SDs below the mean or lower); an additional 16 percent of children are within the at-risk range (between 1.0 and 1.5 SDs below the mean).[16]

Multiple Data Sources Provide a Mixed Picture of Children's Social-Emotional Development

Similar to the assessment of children's language development, we also gathered data from multiple sources to assess children's social-emotional development in different contexts (Box III.3 describes the social-emotional measures and scoring procedures). Table III.6 presents the descriptive statistics for the social-emotional measures at age 3. Early Head Start staff and parents of 3-year-olds completed the Brief Infant Toddler Social Emotional Assessment (BITSEA; Briggs-Gowan and Carter 2006). The BITSEA measures children's emerging social-emotional competence as well as

[14] The standard scores for the national normative sample have a mean of 100 and a standard deviation of 15.

[15] Subsample sizes for the age subgroups range from 52—129. Only one child was 32 months at the time of the ECI assessment; an additional 6 children were 33 months of age. Given these age groups comprised few children, we do not report average estimates at these ages.

[16] During the administration of the ECI assessment, a play partner (a certified Baby FACES assessor) interacts with the child as he or she engages in play with a toy barn set. The play partner's role throughout the duration of the six-minute task is to facilitate the child's play in a manner that elicits the child's communication (while being nondirective and following the child's lead). Given that the play partner was the assessor and not a more familiar adult (such as the child's parent), scores may reflect an underestimate of children's expressive language ability. We asked parents and assessors to rate the shyness of the child as a means of better understanding performance on the ECI. This is further elaborated in the section that follows on children's social-emotional development.

social-emotional and behavior problems. In addition to raw scores, cutoff scores on the two subscales were created to indicate problems for both forms. For the BITSEA Problem subscale, the cutoff point indicates scores at the 75th percentile or higher on national norms. For the Competence subscale, the cutoff point indicates a score at the 15th percentile or lower on national norms, and may suggest a delay in social-emotional competence. Combining the cutoffs in both domains (that is, high levels of problems or low competence) yields a positive screening indicator for the BITSEA. One caveat for the cutoff scores at age 3 is that the BITSEA is designed for use with children up to 36 months of age; however, 58 percent of children included in the Baby FACES age 3 data collection are older than 36 months.[17]

Box III.3. Measures of Child Social Emotional Development

The Brief Infant Toddler Social Emotional Assessment (BITSEA; Briggs-Gowan and Carter 2006) is the screener version of the longer ITSEA, which is designed to detect delays in the acquisition of social-emotional competencies as well as social-emotional and behavior problems in children 12 to 36 months old. The 42-item parent and staff report focuses on the development of competencies (for example, "hugs or feeds dolls or stuffed animals") as well as problem behaviors (for example, "avoids physical contact").

- The 31-item **BITSEA** Problem scale assesses social-emotional and behavioral problems such as aggression, defiance, overactivity, negative emotionality, anxiety, and withdrawal. Higher scores indicate more problems. The 11-item **BITSEA** Competence scale assesses social-emotional abilities such as empathy, prosocial behaviors, and compliance. Lower scores indicate lesser competence. Respondents are asked to rate each item as not true/rarely, somewhat true/sometimes, or very true/often. The BITSEA is available in both English and Spanish and was administered to both parents and teachers/home visitors in the baseline data collection. The raw scores range from 0 to 22 for the Competence domain and 0 to 62 for the Problem domain. At age 3, the scores on the BITSEA Parent Form have unweighted means of 11.0 (SD = 7.6) and 17.8 (SD = 3.3) for the Problem and Competence scales, respectively; the scores on the BITSEA Childcare Provider Form have unweighted means of 6.8 (SD = 6.0) and 16.6 (SD = 3.5) for the Problems and Competence scales, respectively.

[17] Having children older than 3 is not unexpected because the age window allowed children in the 1-year-old Cohort to be up to 15 months of age at enrollment (and therefore, about 39 months old at the last wave of data collection. Children in the Newborn Cohort could have been up to 8 weeks old at study enrollment, and approximately 38 months old at the last wave of data collection.

- We created cutoff scores to indicate either high problems or low competence. Cutoff points were calculated in six-month age bands according to child gender using values established with the national standardization sample on the parent form and child care provider form separately. For the BITSEA problem scale, the cutoff point is at or above the 75th percentile. For the BITSEA competence scale, the cutoff point is at or below the 15th percentile. Scoring in the cutoff range in at least one domain indicates "screening positive" on the BITSEA. At the time of round 3 of Baby FACES data collection, approximately 60 percent of children are 36 months or older, but we still created the cutoff scores using the oldest age band (30-36 months). (Appendix C details the BITSEA norming sample and the psychometric properties observed in this study.)

Bayley Behavioral Rating Scale (BRS; Bayley 1993) measures the child's behavior during child assessment. The BRS is one of the three component scales of the Bayley Scales of Infant Development—Second Edition (Bayley 1993). There are two subscales of the BRS used in Baby FACES:

- **Orientation/Engagement** measures the child's cooperation with the assessor during the assessment, positive affect, and interest in the test materials.
- **Emotional Regulation** measures the child's ability to change tasks and test materials, negative affect, and frustration with tasks during the assessment.

The assessor rates the child's behavior by scoring items on a five-point scale, with 5 indicating more positive behavior (for example, more cooperation and less frustration). Scores are the total of the items in the subscale. Possible scores range from 9 to 45 for Orientation/Engagement and 10 to 50 for Emotional Regulation.

The BRS "non-optimal" cutoff scores indicate raw scores at or below the 10th percentile, and "questionable" cutoff scores indicate raw scores between the 11th and 25th percentile.

Behavior Problem Index (BPI; Zill and Peterson, 1986)**.** The Behavior Problems Index (BPI) was developed by Zill and Peterson to measure child externalizing behavior problems (such as aggression and hyperactivity) and internalizing behavior problems (such as anxiety and depression). Most of the items are from the Achenbach Child Behavior Checklist (Achenbach and Edelbrock, 1981). The National Longitudinal Survey of Youth (NLSY; Center for Human Resource Research, 2009) used the BPI with children 4 years of age or older and the Panel Study of Income Dynamics (PSID) Child Development Supplement used the BPI with children 3 years of age or older. Baby FACES used the BPI in Staff Child Report and Parent Self-Administered Questionnaire, but the items are slightly different than those used in these two studies. The questions ask about specific behaviors that the child may have been exhibited in the past 3 months. Three response categories were used in the questionnaire: (1)"often true," (2) "sometimes true," and (3) "not true". The responses were reverse-coded so that higher scores indicate more behavior problems. In addition to the BPI total score, two subscale scores were created: externalizing and internalizing behaviors. At age 3, both parents and Early Head Start staff rated children's behavior problems on the BPI. The parent-reported BPI have unweighted means of 5.1 (SD =6.0) for the total score and 4.2 (SD = 4.5) and 1.0 (SD = 2.1) for the externalizing and internalizing behavior problems, respectively; the staff reported BPI have unweighted means of 6.6 (SD =8.9) for the total score and 5.4 (SD = 6.4) and 1.6 (SD = 3.4) for the externalizing and internalizing behavior problems, respectively. We also dichotomized the items following the procedure used in NLSY ("often true" or "sometimes true" =1, "not true" = 0). Using this approach, the parent reports have unweighted means of 4.5 (SD = 4.7) for the total score and 3.8 (SD = 3.6) and 0.9 (SD = 1.6) for the externalizing and internalizing behavior problems, respectively; the staff reports have unweighted means of 5.5 (SD = 5.9) and 4.5 (SD = 4.5) and 1.2 (SD = 2.1) for the externalizing and internalizing behavior problems, respectively.

Child Behavior During the Parent-Child Play Assessment. Semi-structured, video-recorded assessments of parent-child play (Two-Bag Task) were coded using the Parent-Child Interaction Rating Scales for the Two-Bag Task (Mathematica Policy Research 2010). Four scales, each ranging from 1 (very low incidence) to 7 (very high incidence), address aspects of children's social and emotional competence:

- **Engagement of Parent** reflects the extent to which the child displays, initiates, and/or maintains interaction with the parent and expresses positive affect toward the parent. Key indicators include approaching or orienting toward the parent, establishing eye contact, engaging the parent in play, and/or positively responding to the parent's play initiations and suggestions. The unweighted mean is 4.77 (SD = 0.94).
- **Sustained Attention** assesses the degree to which the child is involved with the objects presented in the two bags. Indicators include the degree to which the child "focuses in" when playing with an object and the extent to which the child coordinates activities with several objects and/or explores objects in an intentional, focused manner. The unweighted mean is 4.99 (SD = 0.87).
- **Negativity Toward Parent** measures the degree to which child displays expressions of anger, hostility, or disdain toward the parent. Expressions may be overt (for example, forcefully rejecting a toy offered by the parent or pushing the parent away) or covert (for example, hitting or throwing an object in response to a parent's overture). The unweighted mean is 2.21 (SD = 1.25).
- **Enthusiasm** reflects the degree to which the child approaches and participates in the task with vigor, confidence, energy, and eagerness. A child scoring high in enthusiasm expresses an active interest in the task, invests effort in the interaction, shows an appreciation for his or her own successes, and demonstrates a sense of agency and coordination between affect and behavior. The unweighted mean is 4.89 (SD = 0.95).

Staff and parents also completed the adapted Behavior Problems Index (BPI), which was originally developed by Zill and Peterson (1986) to measure child externalizing behavior problems (such as aggression and hyperactivity) and internalizing behavior problems (such as anxiety and depression).

At the end of the one-on-one direct child assessments, the assessor completed the Bayley Behavioral Rating Scale (BRS; Bayley 1993) to evaluate the child's test-taking behaviors during the one-on-one assessment, including Orientation/Engagement and Emotional Regulation. In addition to raw scores on these two subscales, the BRS "non-optimal" cutoff scores on the subscales indicate a raw score at or below the 10th percentile; "questionable" cutoff scores indicate a raw score between the 11th and 25th percentiles on national norms (see Box III.3 for a description of the BRS scores and scoring procedures).

As part of the activities conducted with 3-year-olds and their parents, we also administered a parent-child, play-based assessment in which a number of child behaviors were assessed. Interactions were video-recorded for later coding by a trained team of Mathematica coders using the Parent-Child Interaction Rating Scales for the Two-Bag Assessment (Mathematica Policy Research 2010).[18] Children's social-emotional competence was assessed along four dimensions: (1) the extent to which the child engaged the parent in the play activities; (2) the degree of sustained attention with the provided play materials; (3) expressions of enthusiasm; and (4) displays of negativity toward the parent. Box III.3 provides additional information on each of the observed child behaviors.

[18] Ratings of parent behaviors derived from the Parent-Child Interaction Rating Scales for the Two-Bag Assessment are elaborated below.

Table III.6. Children's Social-Emotional Development at Age 3

Measures	Weighted Means or Percentages (Standard Error)
Parent-Reported BITSEA Raw Score	
Problem domain	10.9 (0.37)
Competence domain	17.8 (0.17)
Staff-Reported BITSEA Raw Score	
Problem domain	6.9 (0.36)
Competence domain	16.7 (0.15)
Parent-Reported BITSEA Cutoff Score	
Problem domain	30.9 (2.39)
Competence domain	16.8 (2.12)
Staff-Reported BITSEA Cutoff Score	
Problem domain	21.7 (1.96)
Competence domain	12.7 (1.59)
Parent-Reported BITSEA Screening Positive (percentage)	39.2 (2.67)
Staff-Reported BITSEA Screening Positive (percentage)	28.2 (2.13)
Parent-Reported BPI	
Externalizing Behaviors	3.8 (0.22)
Internalizing Behaviors	0.9 (0.10)
Total Score	4.6 (0.29)
Staff-Reported BPI	
Externalizing Behaviors	4.6 (0.26)
Internalizing Behaviors	1.3 (0.12)
Total Score	5.6 (0.33)
Assessor-Reported BRS Total Scale Score	
Orientation/engagement	36.0 (0.35)
Emotional regulation	41.2 (0.49)
Assessor-Reported BRS in Non-optimal Range	
Orientation/engagement	15.3 (2.05)
Emotional regulation	10.9 (1.66)
Assessor-Reported BRS in Questionable Range	
Orientation/engagement	12.4 (1.96)
Emotional regulation	10.7 (1.67)
Shyness	
Parent report	
Very shy	2.4 (0.84)
Somewhat shy	27.8 (2.24)
Neither shy nor outgoing	19.0 (2.57)
Somewhat outgoing	19.3 (2.04)
Very outgoing	31.5 (2.65)
Assessor report	
Very shy	3.1 (0.86)
Somewhat shy	16.3 (2.65)
Neither shy nor outgoing	19.7 (2.56)
Somewhat outgoing	29.4 (2.39)
Very outgoing	31.5 (2.66)
Parent-Child Play Assessment	
Engagement of parent	4.7 (0.05)
Sustained attention	5.0 (0.05)
Negativity toward parent	2.3 (0.07)
Enthusiasm	4.9 (0.04)
Sample Size	
Parent SAQ	**443-450**
SCR	**501-507**
Assessor rating	**451-471**
Two-Bag Task	**437**

Sources: Spring 2011 and 2012 Parent Self-Administered Questionnaire (SAQ), Staff-Child Report (SCR), Direct Child Assessment, and Parent-Child Video-Recorded Interaction.

Note: Sample includes both the 1-year-old Cohort and the Newborn Cohort at age 3.

BITSEA = Brief Infant-Toddler Social-Emotional Assessment; BPI = Behavior Problems Index; BRS = Bayley II Behavior Rating Scale.

Parents Reported more Social Emotional Problems for Children than Did Early Head Start Staff Based on the BITSEA

Similar to the findings from the previous rounds of data, parents rated their 3-year-old children as having more behavior problems than did Early Head Start staff. Parents also reported more children as at risk for lacking social competence than did staff. According to parent reports, children's Problem subscale raw score is 10.9 on average, and approximately one-third (31 percent) of children score above the risk cutoff on this subscale. In contrast, staff-reported Problem subscale raw scores average 6.9, and about one-fifth (22 percent) of children score above the cutoff. Staff reports are lower than national norms (25 percent), while parent reports are higher than national norms. Children's Competence subscale raw score is 17.8 and 16.7 on average for parent reports and staff reports, respectively. The proportion of children scoring below the risk cutoff on the Competence subscale is similar according to both parent reports and staff reports (17 and 13 percent, respectively). Based on parent reports, more children have scores that identified them as screening positive for social-emotional issues; Early Head Start staff reports yield a lower identification rate (39 versus 28 percent).

In summary, parents reported more social-emotional problems in 3-year-old Early Head Start children than did Early Head Start staff. Among children who "screen positive" (an indicator of having problems) by either staff or parent report, only 29 percent screen positive according to both. Differences between reporters may reflect the contexts in which children are observed as well as differences in a parent versus a staff member's sense of normative child behavior.

Parents Reported Slightly Fewer Behavior Problems for Children than Did Early Head Start Staff Based on the BPI

In addition to the BITSEA, staff and parents also completed the adapted BPI to report on children's behavior problems. Parents reported fewer externalizing behavior problems as well as fewer internalizing behavior problems in 3-year-olds than did staff (mean = 3.8 versus 4.6 for externalizing behaviors and 0.9 versus 1.3 for internalizing behaviors). As a result, the parent-reported BPI total score was lower than staff reports (mean = 4.6 versus 5.6). Both parent and staff reports on externalizing behavior problems in Baby FACES are lower than parent reports in the Panel Study of Income Dynamics, Child Development Supplement (PSID-CDS), where the mean externalizing score was 7 for 3-year-olds.[19]

Three-Year-Olds' Task Engagement and Emotional Regulation Rated by the Assessors Approach National Norms

According to assessors' rating of children's behavior at the end of direct child assessment, 3-year-old Early Head Start children score 36.0 out of 45 on Orientation/Engagement and 41.2 out of 50 on Emotional Regulation, on average. Approximately 15 percent of children score in the non-optimal range (10th percentile or lower) on Orientation/Engagement; 11 percent score in this same range on Emotional Regulation. About 12 percent of children score in the questionable range (11th to 25th percentile) on Orientation/Engagement, compared to 11 percent on Emotional Regulation.

[19] The items for internalizing behavior problems used in Baby FACES are different than those used in the PSID-CDS; therefore, we are not able to make a comparison on internalizing behavior problems.

Children Display Positive Behaviors in Play Interactions with Their Parents

As observed during the parent-child play interaction, children displayed behaviors in the mid-range (scoring about 5) on engagement, sustained attention, and enthusiasm; expressions of negativity were in the low range and were less frequently observed (2). On average, nearly all children received scores in the mid-to-high range (greater than or equal to 4) on engagement of parent (91 percent), sustained attention with objects during play (95 percent), and expressions of enthusiasm (93 percent). In contrast, only 14 percent of children displayed indicators of negativity in the mid-to-high range. Estimates are comparable to those reported in other large-scale studies with children at this same age, including the EHSREP (ACF 2002a).

Parents and Assessors Rated Similarly on Children's Shyness

At the end of direct child assessment, assessors rated children's shyness during the visit and also asked parents to rate their child's shyness. This rating was a means to help interpret the ECI, in which children might have displayed some degree of inhibition given they were interacting with the assessor (rather than a more familiar adult). Parents and assessors reported a similar proportion of children as very shy (2 versus 3 percent), neither shy nor outgoing (19 versus 20 percent), or very outgoing (32 percent for each).

Parenting and the Home and Neighborhood Environment

At the age 3 data collection, we visited families in their homes, collecting direct child assessment data, video-recordings of parent-child interactions, and observations of the home and neighborhood environment.

Most Children Are Read to or Told Stories at Least Once a Day, Mostly in English

Reading and telling stories to children are important contributors to emergent and later literacy (Raikes et al. 2006). We asked parents of 3-year-olds about the presence of books in the home and how often they read to their child, as well as how often they tell other stories (Table III.7).

Most parents reported that they have several children's books in their homes, most of them in English. Approximately two-thirds of households (64 percent) have more than 25 books for the child, and another one-fifth have 11 to 25 books for the child. Four percent of households have 4 or fewer books. Most households (81 percent) have books in English only. About 17 percent of all households have books in both English and Spanish, and only a few households (1 percent) have books in Spanish only. All children from English-only homes have books in English only. For children from homes where a language other English is spoken, approximately half (52 percent) have books in English only; 44 percent have books in Spanish only; only 4 percent have books in both English and Spanish.

Parents and other adults in the home frequently read to their child and are more likely to read than tell stories. About 58 percent of parents read to their child more than once a day, and an additional one-third (33 percent) read about once daily. Ten percent of children are read to less than once a day. In contrast, about 39 percent of parents tell stories to their child more than once a day; nearly the same proportion (40 percent) tell a story about once a day. Nearly one-fifth (21 percent) of children are told stories less than once a day. Among children who are not read to daily, most (80 percent) are not told stories daily either (Table III.7). Conversely, approximately two-thirds (63 percent) of children who are not told stories daily, are read to daily or more often. There are 8 percent of children who are neither read to nor told stories at least once a day.

Table III.7. Child's Exposure to Reading and Storytelling at Age 3

	Weighted Percentages (Standard Error)
Number of Books in Home	
0	0.0 (0.00)
1 to 4	3.6 (1.14)
5 to 10	12.5 (2.63)
11 to 25	20.1 (3.09)
More than 25	63.8 (3.10)
Language of Books in Home	
English only	81.3 (2.46)
English and Spanish	17.2 (2.42)
Spanish only	1.4 (0.79)
Other only	0.0 (0.00)
English and other	0.0 (0.00)
Frequency Anyone in Household Reads to Child	
More than once a day	57.5 (3.26)
About once a day	32.8 (2.73)
A few times a week	6.7 (1.53)
Once or twice a week	2.1 (0.56)
Less than once a week	1.1 (0.58)
Frequency Anyone in Household Tells Child Stories	
More than once a day	38.9 (2.42)
About once a day	40.3 (2.38)
A few times a week	14.3 (1.93)
Once or twice a week	4.1 (0.89)
Less than once a week	2.5 (0.89)
Percentage of Children Who Are Neither Read to Nor Told Stories at Least Once a Day	7.7 (1.66)
Frequency of Storytelling Among Children Read to Less than Daily	
More than once a day	3.5 (2.77)
About once a day	17.1 (6.61)
A few times a week	42.5 (8.85)
Once or twice a week	21.0 (7.64)
Less than once a week	16.0 (7.38)
Frequency of Reading to Child Among Those Told Stories Less than Daily	
More than once a day	21.4 (5.26)
About once a day	41.5 (5.24)
A few times a week	24.1 (5.78)
Once or twice a week	7.9 (2.31)
Less than once a week	5.1 (2.68)
Language Adult Reads to Child	
English only	77.5 (3.12)
English and Spanish	3.7 (0.79)
Spanish only	18.1 (2.85)
Other only	0.4 (0.28)
English and other	0.4 (0.37)
Language Adult Tells Child Stories	
English only	78.8 (3.03)
English and Spanish	12.5 (2.10)
Spanish only	7.4 (1.33)
Other only	0.0 (0.00)
English and other	1.3 (1.01)
Sample Size	**397-420**

Source: Spring 2011 and 2012 Parent Interview.

Note: Sample includes both the 1-year-old Cohort and the Newborn Cohort at age 3.

Most parents read and tell stories to their children in English, while parents speaking other languages are more likely to use that language alone while reading, but tell stories in both English and the other language. Overall, more than three-quarters of parents read and tell stories to their child in English. About 19 percent of parents read to their child in only Spanish or another non-English language, while about 4 percent read in both English and the other language. When telling stories, however, more parents (about 14 percent) use both English and another language than use only the other language (7 percent).

One-Third of 3-Year-Old Children Speak Two Languages

We also asked parents to report on the languages children speak (Table III.8). Two-thirds of the 3-year-olds (65 percent) enrolled in Early Head Start in spring 2011 and 2012 speak one language, typically English. The remaining children mostly speak two languages. Overall, nearly all children speak some English (97 percent), but only about two-thirds (63 percent) speak it alone. For the remaining one-third of children, most speak some Spanish—about 2 percent speak Spanish only and 29 percent speak both Spanish and English; 4 percent speak another language and English (less than 1 percent speak another language only).

We asked parents to identify the child's first language as well as the languages the child speaks most often. Three-quarters of parents said that their child's first language is English. Of the remaining one-quarter of children, most (23 percent) have Spanish as their first language and less than 1 percent have both English and Spanish as their first language; 1 percent have other language as their first language and less than 1 percent have both English and other language as their first language. For the majority of the 3-year-olds, the language the child speaks most often is the same as the child's first language, but about 3 percent of parents said that their child does not speak their first language most often. Of those, about three-quarters are children whose first language is Spanish or other language but who speak English most often.

Table III.8. Languages Spoken by Child at Age 3

	Weighted Percentages (Standard Error)
Number of Languages Child Can Speak	
One	65.2 (3.77)
Two or more	34.8 (3.77)
Child's Language	
English	
English only	63.1 (3.83)
Spanish	
Spanish only	1.9 (0.44)
Spanish and English	29.4 (3.68)
Other	
Other only	0.2 (0.22)
English and other	3.8 (1.29)
Child's First Language	
English	74.5 (3.17)
Spanish	23.2 (3.20)
English and Spanish	0.4 (0.29)
Other	1.2 (0.57)
English and other	0.6 (0.49)
Child's First Language is Different from Language Child Speaks Most Often	2.7 (0.87)
Sample Size	**420-421**

Source: Spring 2011 and 2012 Parent Interview.

Note: Sample includes both the 1-year-old Cohort and the Newborn Cohort at age 3.

Most Home Environments Are Emotionally Supportive and Cognitively Stimulating

During in-person home visits to the families of 3-year-olds, interviewers completed the Home Observation for Measurement of the Environment (HOME) inventory (Caldwell and Bradley 2003), based on their observation in the home with the child's parents while the child is present (Table III.9). The HOME inventory measures the quality of stimulation and support available to a child in the home environment, and has been found to be predictive of many later child outcomes (Bradley 2006; Bradley and Corwyn 2007; Bradley et al. 2001; see Box III.4 for a description of the HOME scores and scoring procedures).

Table III.9. Home and Neighborhood Environment at Age 3

	Weighted Means (Standard Error)
HOME	
Parental Warmth	2.5 (0.04)
Parental Lack of Hostility	2.4 (0.10)
Support of Cognitive and Learning Environment	11.4 (0.10)
Internal physical environment	2.3 (0.08)
HOME total score	18.7 (0.20)
Sample Size	**350-449**

Source: Spring 2011 and 2012 Parent SAQ.

Note: Sample includes both the 1-year-old Cohort and the Newborn Cohort at age 3.

HOME = Home Observation for Measurement of the Environment (HOME); FES = Family Environment Scale.

High levels of emotional support. Parents of 3-year-old children exhibited high levels of warmth toward their children, as observed by the assessors during home visits. The mean score on the HOME Parental Warmth subscale is 2.5 out of 3, which is in line with the mean score of 2.6 found in the EHSREP. Nearly two-thirds (64 percent, standard error [SE] = 2.38) of parents have a score of 3 on this subscale in the Baby FACES sample (not shown).

Lack of hostility. Assessors rated parents of 3-year-olds as generally absent of harsh or punitive parenting behaviors. The mean score on the Lack of Hostility subscale is 2.4 out of a possible 3. Approximately three-quarters of parents have a score of 3.

Support of cognitive and learning environment. Parents of 3-year-old children provided support for children's cognitive development and learning by providing cognitive stimulation during interaction with the child, helping the child to learn shapes and sizes, alphabet, colors, and numbers, frequent reading to the child, and providing cognitively stimulating books and materials in the home. The mean score on the Support of Cognitive and Learning Environment subscale is 11.4 out of 13. This is slightly higher than the results from the EHSREP, where the mean score was 10.6 when the children were 3 years old.

Internal physical environment of the home. Assessors rated 3-year-old children's homes as generally clean and organized, with a mean of 2.3 out of 3 on this subscale. More than half (52 percent) of parents have a score of 3 on this subscale.

Total HOME score. Overall, the total HOME scores average 18.7 out of 22, suggesting that 3-year-old children live in home environments that have adequate emotional support and cognitive and language stimulation.

Box III.4. Measures of Home and Neighborhood Environment

- **The Family Environment Scale, Family Conflict Subscale** (FES; Moos and Moos 2002) was designed to measure the social and environmental characteristics of families, including family relationships, emphases on aspects of personal development that families can support, and maintenance of the family system. The Family Conflict subscale measures the extent to which the open expression of anger and aggression and conflict-filled interactions are characteristic of the family. Parents rated each of five items on a four-point scale, in which 4 indicates strong agreement with statements such as "We fight a lot" and "We sometimes hit each other." The subscale score is then the mean of the five individual item scores. For the Baby FACES sample, however, we removed one item that had a low correlation with the rest of the items in the scale and therefore reduced the overall alpha of the measure: "We hardly ever lose our tempers." The unweighted mean using the four other items for Family Conflict Subscale as reported by Baby FACES parents is 1.4 (SD = 0.4) at round 3 of data collection.

- **Exposure to Violence** measures how many violent incidents (out of four) a child has observed or been a victim of in his or her lifetime. Items come from the Infant-Toddler Social and Emotional Assessment, in which parents are asked to respond yes or no to questions whether a child has in his or her neighborhood: seen violence in their neighborhood, been a victim of violence, seen someone use a weapon to hurt or threaten a family member, or seen someone hit, push or kick a family member (Carter and Briggs-Gowan 2000). The unweighted mean is 0.1 (SD = 0.5) at round 3 of data collection.

- **Home Observation for Measurement of the Environment** (HOME) Home Observation for Measurement of the Environment (HOME) measures the quality of stimulation and support available to a child in the home environment (Caldwell and Bradley 2003). The age-3 assessment was based on the HOME-Short Form inventory, Preschool version, and the neighborhood rating items were from the Project on Human Development in Chicago Neighborhood (PHDCN). Information needed to score the inventory is obtained through a combination of parent self-administered questionnaire and observation conducted in the home with the child's parent while the child is present. We derived five subscales from this assessment, as well as the total score:

  - **Emotional Responsivity** measures responsive and supportive parenting behavior observed by the interviewer during the home visit. Items in this subscale are based entirely on interviewer observations of the parent and child during the interview, and include such items as whether the mother praised the child, whether she expressed warmth and affection toward the child, and whether her voice conveys positive feelings toward child. Scores can range from 0, if none of the positive behaviors were observed, to 3, if all of the behaviors were observed. The unweighted mean is 2.5 (SD = 0.8) for the Emotional Responsivity subscale.

  - **Support of Language and Learning** measures the breadth and quality of the mother's speech and verbal responses to the child during the home visit, as rated by the interviewer; whether the parent encourages the child to learn shapes, colors, numbers, and the alphabet; the presence of books, toys, and games accessible to the child; and whether the parent reads to the child several times per week. Items are obtained by a combination of parent report and interviewer observation. Possible scores range from 0 to 13. The unweighted means for Support of Language and Learning subscale is 11.4 (SD = 1.7).

  - **Absence of Punitive Interactions** measures lack of harsh or punitive parenting behavior observed during the home interview. Items in this subscale are based entirely on interviewer observations of the parent and child during the interview, and include whether the parent scolded the child, physically restrained the child, or slapped or spanked the child. Items are scored 1 if the parent did not engage in particular harsh or punitive behaviors during the home visit. Scores can range from 0 to 3. Higher scores on this outcome measure imply less negative parenting behavior. The unweighted mean is 2.5 (SD=1.0) for this subscale.

- **Internal Physical Environment** measures the cleanliness, organization, and warmth of the home environment. Items in this subscale are based entirely on interviewer observations during the interview. Scores can range from 0 to 3. The unweighted mean for this subscale is 2.3 (SD = 0.9).
- **Total Score** measures the cognitive stimulation and emotional support provided by the parent in the home environment. The total includes 22 items used in the four subscales. The maximum potential score is 22. The unweighted mean of the total score is 18.9 (SD = 2.8).
- **External Environment (not included in the HOME total)** measures the physical and social environment of the face-block. Items in this subscale are based entirely on interviewer observations of the neighborhood, and include such items as general condition of most of the housing units, garbage in the street or on the sidewalk, volume of traffic, and people arguing or fighting in the street. The items are recoded as 1 (yes) or 0 (no), and then summed up. Scores can range from 0 to 8. The unweighted mean for this measure is 6.5 (SD=1.6).

- **Neighborhood Disorder** measures the physical and social environment of the face-block where the family lives. Items in this subscale are based entirely on interviewer observations of the neighborhood and include such items as general condition of housing units, garbage in the street or on the sidewalk, volume of traffic, and people arguing or fighting in the street. The scale score is the mean of the item z-scores. Higher scores indicate higher levels of disorder.

Children Generally Experience Supportive, Positive Interactions with Their Parents

As noted above, we also administered an 8-minute parent-child play task in which a number of child and parent behaviors were assessed. For the assessment of parent behaviors, video-recorded play sessions were coded by a trained team of Mathematica coders using two coding schemes: the Parent-Child Interaction Rating Scales for the Two-Bag Assessment (Mathematica Policy Research 2010) and an adaptation of the Parenting Interactions with Children: Checklist of Observations Linked to Outcomes (PICCOLO; Roggman et al. 2009). Collectively, the coding schemes assess positive and negative parenting behaviors that are meaningfully linked to children's developmental outcomes. Box III.3 provides additional information on each of the child behaviors that are of focus in the Two-Bag and Box III.5 provides information on the parent behaviors in the Two-Bag and PICCOLO coding schemes, including the unweighted means for each of the scales.

According to ratings on the Parent-Child Interaction Rating Scales (Table III.10), parents of 3-year-old children received scores in the mid-range on sensitivity (4.5), positive regard (4.3), stimulation of cognitive development (4.4), and relationship quality (4.5). Parental sensitivity, positive regard, and relationship quality were highly intercorrelated, and were combined into a single composite score (synchronicity), with comparable average ratings in the mid-range (4.4). Most parents received scores in the mid to high range (greater than or equal to 4) on sensitivity (84 percent), positive regard (81 percent), stimulation of cognitive development (85 percent), and quality of the relationship (83 percent). Conversely, negative parenting behaviors during the play-based assessment, including negative regard (2.3), intrusiveness (3.0), detachment (2.4), and dissolution of boundaries (2.5) were in the low range and were less frequently observed. A majority of parents received scores in the low range (less than or equal to 3) on intrusiveness (71 percent), and at least two-thirds of parents displayed low-level indicators of detachment (85 percent), negative regard (82 percent), and dissolution of boundaries (81 percent). Average scores among parents of 3-year-old children in Baby FACES are slightly more favorable when compared to estimates observed at age 2. In addition, average ratings

from Baby FACES are within the same range as those reported in other large-scale studies with children at this same age, including the EHSREP (ACF 2002b).

Table III.10. Parent-Child Play Assessment Parent Scales at Age 3

Scales	Weighted Mean (Standard Error)
Parent-Child Interaction Rating Scales	
Sensitivity	4.5 (0.07)
Positive Regard	4.3 (0.07)
Stimulation of Cognitive Development	4.4 (0.06)
Quality of Relationship	4.5 (0.06)
Synchronicity	4.4 (0.05)
Negative Regard	2.3 (0.07)
Intrusiveness	3.0 (0.07)
Detachment	2.4 (0.06)
Boundary Dissolution	2.5 (0.07)
PICCOLO Total Score	1.5 (0.02)
Affection	1.7 (0.01)
Responsiveness	1.5 (0.02)
Encouragement	1.5 (0.02)
Teaching	1.4 (0.02)
Sample Size	**437**

Source: Spring 2011 and 2012 Parent-Child Video-Recorded Interaction, Two-Bag Task.

Note: Sample includes both the 1-year-old Cohort and the Newborn Cohort at age 3.

PICCOLO=Parenting Interactions with Children: Checklist of Observations Linked to Outcomes.

We also assessed positive parenting behaviors using the PICCOLO, an observational instrument designed to measure developmentally appropriate parenting along four domains: affection, responsiveness, encouragement, and teaching. Ratings averaged 1.7 for affection, 1.5 for responsiveness, 1.5 for encouragement, 1.4 for teaching, and 1.5 for the overall score (out of a possible 0 to 2 range). These average scores are similar to those reported in other studies with parents of children at this same age (Cook and Roggman 2009). Based on recommended age-based scoring rubrics provided by the developers for each of the four PICCOLO domain total scores (Roggman et al. 2009), a majority of parents display behaviors in the high range. Specifically, 89 percent of families score in the high range on affection (total scores greater than or equal to 10 out of 14); 74 percent on encouragement (total scores greater than or equal to 10 out of 14); and 91 percent on teaching (total scores greater than or equal to 8 out of 16). A smaller proportion (53 percent) of parents scored in the high range on responsiveness (total scores greater than or equal to 11 out of 14).

Children Are Rarely Exposed to Violence or Conflict in Their Homes

To measure a child's exposure to violence, we used items from the Infant-Toddler Social and Emotional Assessment (ITSEA; Carter and Briggs-Gowan 2000), asking parents about their child's exposure to three types of violence: (1) violence in the neighborhood, (2) a weapon used to threaten or hurt a family member, and (3) someone hitting, pushing, or kicking a family member. Additionally, we asked parents whether their child had been a victim of violence in the neighborhood (Table III.11).

The majority of 3-year-olds in Early Head Start (88 percent) have not been exposed to any of the three types of violence or been a victim of violence in the neighborhood. About 9 percent of children have encountered one of the four types of violence, and 3 percent have witnessed two or more. Overall, parents most commonly reported that their child had witnessed someone hitting, pushing, or kicking a family member (10 percent, SE=1.51), followed by the child having seen violence in his or

her neighborhood (3 percent, SE= 0.78), or seen a weapon used against a family member (2 percent, SE=0.86). Less than 1 percent (SE=0.40) of children have been victims of violence.

Table III.11. Child's Exposure to Violence and Neighborhood Disorder at Age 3

	Weighted Means or Percentages (Standard Error)
Number of Violence Types to Which Child Has Been Exposed[a]	
Zero	88.3 (1.78)
One	8.8 (1.68)
Two or more	2.8 (0.81)
FES-Family Conflict	1.4 (0.03)
External Environment	6.6 (0.11)
Neighborhood Disorder[b]	-0.1 (0.04)
Sample Size	**377-402**

Sources: Spring 2011 and 2012 Parent Interview and Parent SAQ.

Note: Sample includes both the 1-year-old Cohort and the Newborn Cohort at age 3.

[a]Number is the sum of yes responses to four questions regarding acts of violence: (1) whether a child has ever seen violence in their neighborhood; (2) whether a child has seen someone use a weapon to threaten or hurt a family member, (3) whether a child has seen someone hit, push, or kick a family member, and (4) whether the child has been a victim of violence in the neighborhood.
[b]The scale score is the mean of the item *z*-scores. Higher scores indicate more disorder in the neighborhood.

Overall, parents of 3-year-olds also report low levels of household conflict. We asked parents to rate their level of agreement with statements about conflict among members of their household, using the 4-point Family Conflict subscale of the Family Environment Scale (FES) (see Box III.4 for a description of the FES scores and scoring procedures). The subscale mean measures the average level of agreement that parents reported for statements on the open expression of anger and aggression and conflict-filled interactions within their family; higher scores indicate a higher level of agreement to statements about incidence of anger in the home. On average, parents of 3-year-olds report a low level of household conflict, with an average score of 1.4. Less than 7 percent (SE=1.47) of parents had an FES score above 2. Of children who have been exposed to any violence (N = 45), about 21 percent (SE=7.88) also have parents who report an FES score above 2.

Many Families of 3-Year-Old Children Live in Neighborhoods with Poor Conditions

Assessors rated the physical and social environment of the face-block (generally, the block on which the families lives) upon exiting the in-person home visit. The neighborhood rating items are drawn from the Project on Human Development in Chicago Neighborhoods (PHDCN; Sampson 2012), and the ratings are based entirely on assessor observations of the neighborhood. The scores are a count of negative neighborhood conditions (such as garbage and/or drug paraphernalia in the street or on the sidewalk and people outside arguing or fighting), with a possible highest score of 8. Higher scores indicate more disorder in the neighborhood. On average, the assessors rated the neighborhoods as being in poor condition, with a score of 6.6 out of 8. Nearly two-thirds (63 percent, SE= 3.39) of the families have a score of 7 or 8. A neighborhood scoring an 8 would have some or all of the following characteristics: badly deteriorated housing units or buildings; streets with potholes, garbage, litter, and trash "just about everywhere"; drug paraphernalia, cigarette butts, condoms, or beer or liquor containers or packages "just about everywhere"; heavy traffic that backs up at traffic lights; several adults or teenagers on the street arguing or otherwise behaving in a hostile manner; and a general atmosphere that had the observer fearing for his or her personal safety.

Box III.5. Parent Scales from the Parent-Child Play Assessment

- **The Parent-Child Interaction Rating Scales for the Two-Bag Task** consists of 12 scales that assess a range of child and parent behaviors. A total of 8 scales address both positive and negative parenting behaviors. Each behavior is rated along a 7-point scale, ranging from a very low incidence to a very high incidence of the behavior. We created a composite parenting score, synchronicity (Cronbach's alpha=0.85), by computing a mean score derived from scores on parental sensitivity, positive regard, and relationship quality—all of which were highly and significantly correlated (ranging from 0.57 to 0.78).
  - **Sensitivity** measures the degree to which the parent responds to the child's cues (such as gestures, expressions, and signals) during times of both distress and nondistress. The defining characteristic of maternal sensitivity is its child-centered focus, which includes "tuning in" to the child; manifesting an awareness of the child's needs, moods, interests and capabilities; and being flexible in supporting and responding to the child's emerging need for autonomy.
  - **Positive Regard** assesses the parent's expression of love, respect, and/or admiration for the child. Key indicators include verbal or physical praising of the child's efforts and successes, words of encouragement or support, and nonverbal expressions of affection. Additional exemplars include clear enjoyment of the child, displays of interest in his or her play, and expressions of concern and/or empathy for the child's distress.
  - **Stimulation of Cognitive Development** measures the quality and quantity of the parent's efforts to enhance the child's perceptual, cognitive, and linguistic development. Key features include attempts to stimulate higher levels of mastery and sophistication matched to or slightly above the child's developmental level and interest, and the use of complex and varied language. This scale weights heavily opportunities that encourage and/or facilitate pretend play.
  - **Quality of the Relationship** assesses the degree of affective sharing and reciprocity between the parent and the child. Quality interactions are characterized by a sense of emotional relatedness and mutual engagement, contingent responding, and displays of affective and/or verbal sharing.
  - **Negative Regard** reflects the parent's expression of discontent with, anger toward, and disapproval and/or overt rejection of the child. This dissatisfaction may be manifested verbally (for example, derogatory words or disregard toward the child) or physically (for example, threatening posture or physical roughness). Additional indicators of negativity include an underlying sense of frustration with the child and abrupt or curt responses to the child's bids for attention.
  - **Intrusiveness** reflects the extent to which the parent exerts control over the child rather than acts in a way that acknowledges and respects the child's perspective. Intrusive interactions are adult-centered and involve imposing the parent's agenda on the child despite signals that a different activity, level, or pace of play is needed. Expressions of intrusiveness can be physical or verbal.
  - **Detachment** measures the parent's lack of awareness, attention to, and engagement with the child. Key indicators include a lack of emotional responsiveness to the child's bids for attention, interacting with the child in a perfunctory or indifferent manner, or responding in a way that is not contingent on, or "out of sync" with, the child's affect, actions, or vocalizations.
  - **Physical and/or Psychological Dissolution of Boundaries** refers to the extent to which the parent fails to maintain an appropriate parental role in his or her interaction with the child. Interactions characterized by boundary dissolution lack clear distinctions between the parent and the child, as demonstrated in the parent's inability to provide firm directives, set appropriate limits, and/or provide the child with clear expectations for behavior. Displays of boundary dissolution can be psychological or manifest in inappropriate physical behaviors.
- **The Parenting Interactions with Children: Checklist of Observations Linked to Outcomes** (PICCOLO; Cook and Roggman 2009; Roggman et al. 2009) is an observational instrument designed to measure positive parenting along four domains known to support children's early development: (1) affection, (2) responsiveness, (3) encouragement, and (4) teaching. Twenty-nine behaviors are rated on

a three-point scale, ranging from 0 (absent) to 2 (clearly evident). Behaviors that are infrequently or barely observed are indicated by a score of 1. The domains of affection, responsiveness, and encouragement each comprise seven items; the teaching scale consists of eight items. Mean scores are derived by averaging across the component items. A total score is also computed by averaging the domain scores. Appendix C further details inter-rater reliability estimates and scoring procedures.

- **Affection** measures the extent to which the parent displays warmth, physical closeness, and positive expressions toward the child. Items include the degree to which the parent speaks in a warm tone of voice, smiles at the child, praises the child, maintains close physical proximity, displays positive expressions of affect, actively engages the child in the interaction, and demonstrates emotional support.
- **Responsiveness** assesses the frequency with which the parent responds to the child's cues, emotions, vocalizations, interests, and behaviors. Items include the extent to which the parent attends to the child's actions, adjusts the activity, level, or pace of play as needed to align with the child's interests and/or needs, demonstrates flexibility in supporting the child's interests, follows the child's lead, responds to the child's displays of emotion, visually orients toward the child in response to the child's vocalizations, and responds verbally to the child's vocalizations.
- **Encouragement** reflects the degree to which the parent actively supports the child's exploration, effort, skills, initiative, curiosity, creativity, and play. Items include the extent to which the parent provides the child with ample time to respond after offering a suggestion, encourages the child to explore the play materials, supports the child's choice of activity, supports the child's need for autonomy, verbally encourages the child's efforts, scaffolds the child's play, and shows enthusiasm in response to the child's efforts.
- **Teaching** assesses the degree to which the parent engages in shared conversation and play, provides cognitive stimulation, and extends the child's verbalizations. Items include the frequency with which the parent provides explanations, suggests activities that extend the child's actions, repeats or expands on the child's vocalizations, labels an object or action, engages in pretend play, performs activities in an ordered sequence of steps, describes features or characteristics of objects, and asks the child for information (for example, by posing questions).

Parent-Child Play Assessment Parent Scale Scores, Unweighted

Scales	Unweighted Mean	Standard Deviation
Parent-Child Interaction Rating Scales		
Sensitivity	4.5	1.02
Positive regard	4.3	1.02
Stimulation of cognitive development	4.5	0.96
Quality of relationship	4.5	1.07
Synchronicity	4.4	0.91
Negative regard	2.4	1.13
Intrusiveness	2.9	1.24
Detachment	2.4	1.01
Boundary dissolution	2.5	1.15
PICCOLO Total Score	1.5	0.24
Affection	1.7	0.26
Responsiveness	1.5	0.32
Encouragement	1.5	0.34
Teaching	1.4	0.31
Sample Size	**444**	

Source: Spring 2011 and 2012 Parent-Child Video Interaction, Two-Bag Task.

PICCOLO=Parenting Interactions with Children: Checklist of Observations Linked to Outcomes.

Summary of Key Findings

- Most 3-year-old children in Early Head Start maintain physical well-being and have access to health care.
- On average, 3-year-old Early Head Start children scored similarly to national norms in general development.
- Early Head Start children demonstrate good language skills on some measures at age 3, but lag the normative sample on others.
  - Parents reported better expressive language development in children than did Early Head Start staff.
  - Parents reported that most children demonstrate good language comprehension and production skills.
  - Direct child assessment suggests that Early Head Start children's auditory comprehension as measured by the PLS-4 is approaching national norms, while receptive vocabulary as measured by the PPVT-4 and expressive language skills as measured by the ECI still lag the normative sample.
- Multiple data sources provide a mixed picture of children's social-emotional development.
  - Parents reported more social-emotional problems for children than did Early Head Start staff based on the BITSEA; however, parents reported slightly fewer behavior problems for children than did Early Head Start staff based on the BPI.
  - Three-year-olds' task engagement and emotional regulation rated by the assessors are approaching national norms.
  - Children display positive behaviors in play interactions with their parents.
- Parents are generally supportive of their children's development; although many children live in poor neighborhood environments.
  - Most children are read to or told stories at least once a day, mostly in English.
  - One-third of 3-year-old children speak two languages.
  - Most home environments are emotionally supportive and cognitively stimulating.
  - On average, parents display positive behaviors during interactions with their children; negative parenting behaviors are less frequently observed.
  - Children are rarely exposed to violence or conflict in their homes.
  - Many families of 3-year-old children live in neighborhoods with poor conditions.

IV. CHILD AND FAMILY OUTCOMES IMPROVED OVER THE EARLY HEAD START YEARS

As a two-generation program, Early Head Start is expected to bring about positive family and child outcomes over the course of program enrollment, with the ultimate goal being improvements in children's competence. Chapter III of this report, together with the prior two Baby FACES reports, delineated the functioning of children and families at each point of data collection. In this chapter, we describe the trajectories of growth/change for child and family outcomes during the period of enrollment in Early Head Start using longitudinal data. The following research questions are addressed:

- How do child and family outcomes change over time in Early Head Start?
- Do the trajectories of growth/change differ by child, family, and program characteristics?

To answer these questions, we use multilevel growth curve analysis to investigate how the child and family outcomes change over time and explore the predictors of the change trajectories. Growth curve analysis allows us to capitalize on the longitudinal nature of the data. More specifically, rather than simply examining the average levels of particular outcomes at each time point, growth curve analysis allows us to model continuous change over time. In addition, we can get a better understanding of how children's or families' progress or growth over time may differ depending on child, family, and program characteristics. In these analyses, we focus on changes in child language and social-emotional development and parent mental health and parenting stress over time in Early Head Start. We include child and family characteristics (child gender, race/ethnicity[20], DLL status, child health, family income-to-needs ratio, maternal demographic risk, and parent psychological risk) and program characteristics (program approach) in the model to examine how these characteristics are related to individual change in the outcome measures during the program years. As we noted in Chapter II, these are nonexperimental analyses of children's development and change in parent outcomes, and it is possible that changes in outcomes over time are related to changes in the composition of the sample.[21]

Box IV.1. Growth Curve Model and Covariates

We are able to conduct growth curve analysis to examine change over time in outcomes for which we have data at three or four time points. These curves describe the average level of the outcomes, rate of growth or change on the outcome of interest and how it changes over time (that is, the acceleration or deceleration of the change rate if you have enough time points to model a nonlinear effect). Because children and families in the same program are likely to be more similar to each other than to children and families in other programs, using multilevel modeling to take the nested data structure into account will improve the precision of estimates of the associations between change in child and family outcomes and child and family and program characteristics. Therefore, we estimated a three-level model using the Hierarchical Linear modeling (HLM) software (Raudenbush & Bryk, 2002) for each child and family outcome to characterize the variation in the growth or change in the outcome: within individuals over time (level 1), between individuals within programs (level 2), and between programs (level 3). (Appendix D provides further details of this modeling approach.)

[20] Race/ethnicity is not included in the models of the CDI because DLL status is confounded with race/ethnicity.

[21] As a sensitivity test, we also conducted growth curve analysis using children with data at all three time points. The results indicate that the shapes of the curves are similar in both sets of analysis, making it less of a concern that more disadvantaged children and families dropped out. See Appendix D for details about the sensitivity analyses.

The level-1 model estimates growth/change over time within individuals. It demonstrates the trajectory over time for each of the outcomes—whether it is linear or nonlinear. Given that we have three to four time points available for the data, we model both a linear and a quadratic component in the level-1 model and drop the quadratic component if it is not significant. The child outcomes are modeled as a function of the child's age in months at each wave of data collection (and thus control for different intervals of time between waves for different children). The parent outcomes are modeled as a function of data collection waves.

The level-2 model estimates variation in growth or change between individuals within programs. By including the child and family characteristics in the model, we examine which ones significantly predict the change trajectories of the outcomes. For child outcomes, we include the following covariates: child gender, race/ethnicity, DLL status, child health (low birth weight, premature birth, child with good or excellent health), family income-to-needs ratio, maternal demographic risk, and parent psychological risk. For parent depressive symptoms and parenting stress, we examine child gender, race/ethnicity, DLL status, family income-to-needs ratio, and maternal demographic risk.

Level-3 models variation between programs. The purpose of the program-level model is to examine how program characteristics are related to individual change in the outcome measures. Specifically, we focus on program service approach.

Children Show Growth in Language and Social-Emotional Domains

In Baby FACES, we draw on data from multiple sources to capture children's language and social-emotional development across different contexts. However, only measures that are available across three rounds of data collection are appropriate candidates for growth curve analysis. Very few measures exist that can assess children from age 1 to age 3. Baby FACES used several of these measures including: Early Head Start teacher or home visitor reports on children's vocabulary comprehension and production using the MacArthur-Bates Communicative Development Inventories (CDI)—Short Form (Fenson et al. 2000) (see Box VI.2 for information about the CDI IRT scores; Box III.2 provides general information about the CDI); the Ages & Stages Questionnaires, Third Edition (ASQ-3; Squires et al. 2009) Communication scores (parent reports); and Early Head Start staff and parent reports of the Brief Infant Toddler Social Emotional Assessment (BITSEA; Briggs-Gowan and Carter 2006). As a screening tool, the ASQ-3 Communication raw scores have a ceiling problem and are not suitable for examining growth. For the BITSEA, the Baby FACES Technical Work Group recommended we focus on parent reports because a teacher or home visitor's time with a child is far less than a parent's and staff, particularly home visitors, may not have the opportunity to see children at their most challenging moments. Moreover, different Early Head Start staff rated children across rounds of data collection. Thus, for child outcomes, we examine children's language development using staff reports on the CDI and social-emotional development using parent reports on the BITSEA. Based on the literature we expect that children's vocabulary skills will increase over time in Early Head Start but the rate of growth will slow down over time (Fenson et al. 2007). We also expect that children's social competence will increase and problem behaviors will decrease over time in Early Head Start, but it is not clear whether to expect linear or nonlinear change.

Table IV.1 shows the descriptive statistics of the CDI IRT scores (see Box VI.2 and Appendix D for a description for the procedure of creating the CDI IRT scores) and BITSEA raw scores at each point of data collection. As shown in Table IV.1, there is a general trend increase in the CDI scores, or vocabulary skills, and BITSEA Competence scores over time. Problem behavior (BITSEA Problem) scores increase from age 1 to age 2, but decrease from age 2 to age 3. Although the means for vocabulary skills and social competence illustrate general increasing trends, growth curve models can examine whether these trends are significant.

Table IV.1. Child and Family Outcomes Improve Over Time (Unweighted)

	Age 0[a]		Age 1		Age 2		Age 3	
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
Child Outcomes								
English CDI IRT Score	NA	NA	842	29.02 (11.17)	651	52.36 (12.61)	517	61.49 (10.24)
Spanish CDI IRT Score	NA	NA	155	33.08 (9.60)	133	50.97 (12.38)	101	60.27 (13.00)
Parent-Reported BITSEA Raw Score								
Competence	NA	NA	771	16.14 (3.41)	589	16.97 (3.44)	452	17.80 (3.22)
Problem	NA	NA	777	10.48(6.19)	589	12.32 (7.51)	450	11.03 (7.50)
Parent Outcomes								
CESD-SF Raw Score	117	6.49 (6.29)	752	5.17 (5.55)	517	4.05 (5.83)	438	3.99 (5.96)
PSI-SF Raw Score								
Parental Distress	NA	NA	749	10.77 (4.69)	595	10.40 (4.75)	439	9.63 (4.71)
Parent-Child Dysfunctional Interaction	NA	NA	748	8.63 (4.00)	594	8.58 (4.39)	439	8.65 (4.69)

Source: Spring 2009 to spring 2012 Staff Child Report (SCR) and Parent Interview.

Note: Sample includes both the 1-year-old Cohort and Newborn Cohort children.

[a]Assessed for pregnant women in the Newborn Cohort.

SD = Standard deviation; CDI= MacArthur-Bates Communicative Development Inventories; BITSEA = Brief Infant-Toddler Social-Emotional Assessment; PSI = Parenting Stress Index; CESD-SF = Center for Epidemiologic Studies Depression Scale Short Form; NA = Not applicable.

Box IV.2. English and Spanish CDI IRT Scores

Baby FACES uses the CDI to assess children's language abilities across time. For all children in the sample, Early Head Start staff rated children's vocabulary comprehension and production using the CDI Infant Form at age 1, the CDI Toddler Form at age 2 and the CDI-III at age 3. In addition, those who reported working with children who understood Spanish (and who themselves understood Spanish) were asked to complete the CDI forms in Spanish at each wave of data collection (see Chapter III of this report and previous two reports at baseline and age 2 [Vogel et al. 2011; Vogel et al. 2015] for descriptions of the different CDI forms). For each word on the forms, the staff rated the child on a 3-point scale: does not understand (0), understands (1), and understands and says (2). We used item response theory (IRT) analysis to scale the three CDI forms together to create the IRT scores. We did this separately for English and Spanish forms and used the IRT scores in the growth curve analysis.

Children's Receptive and Expressive Vocabulary Skills Increase over the Early Head Start Years

The results from the three-level growth curve analysis indicate that children's English vocabulary skills grow in a non-linear fashion during their time in Early Head Start; specifically vocabulary growth is increasing but slowing down over time—(Table VI.2 and Figure IV.1). The shape of the growth trajectory is similar to the cross-sectional age trend for children in the English CDI normative sample (Fenson et al. 2007).

Table IV.2. Predictors of Children's English Vocabulary Growth Trajectory

	Model 1	Model 2	Model 3
	Coefficient (Standard Error)	Coefficient (Standard Error)	Coefficient (Standard Error)
Mean CDI IRT Score	44.90 (0.80)***	48.67 (1.42)***	48.44 (1.43)***
Program service approach			
Center-based			1.25 (1.96)
Home-based			1.02 (1.11)
Multiple-approach (referent)			
Male		-1.02 (0.89)	-0.99 (0.87)
DLL status		-2.84 (1.28)*	-2.83 (1.27)*
Family income-to-needs ratio		0.06 (0.16)	0.06 (0.16)
Maternal demographic risk[a]			
Lower risk (referent)			
Moderate risk		-1.63 (0.95)+	-1.65 (0.95)+
High risk		-2.69 (1.24)*	-2.72 (1.23)*
Parent psychological risk[b]			
No risk (referent)			
One risk factor		-1.81 (1.10)	-1.81 (1.10)
Two or more risk factors		-1.31 (1.80)	-1.28 (1.80)
Low birth weight		-2.99 (2.12)	-3.00 (2.12)
Premature birth		-2.44 (2.50)	-2.44 (2.51)
Child with good or excellent health		0.13 (1.18)	0.11 (1.19)
Age Slope	1.47 (0.04)***	1.59 (0.08)***	1.59 (0.08)***
Male		-0.04 (0.04)	-0.04 (0.04)
DLL status		-0.06 (0.05)	-0.06 (0.05)
Family income-to-needs ratio		0.00 (0.01)	0.00 (0.01)
Maternal demographic risk			
Low risk (referent)			
Moderate risk		-0.06 (0.08)	-0.06 (0.08)
High risk		-0.04 (0.07)	-0.04 (0.07)
Parent psychological risk			
No risk (referent)			
One risk factor		-0.05 (0.07)	-0.05 (0.07)
Two or more risk factors		-0.09 (0.13)	-0.09 (0.13)
Low birth weight		-0.06 (0.14)	-0.06 (0.14)
Pre term birth		0.05 (0.14)	0.05 (0.14)
Child with good or excellent health		-0.03 (0.08)	-0.03 (0.08)
Squared Age	-0.025 (0.003)*	-0.025 (0.004)*	-0.025 (0.004)*

Source: Spring 2009 to spring 2012 Staff Child Report (SCR) and Parent Interview.

Note: Sample includes both the 1-year-old Cohort and Newborn Cohort children.

[a]Maternal demographic risk is the sum of five maternal characteristics collected at baseline: (1) single, (2) a teenager at first birth, (3) lacking a high school or equivalent credential, (4) receiving public assistance, or (5) not employed or in school or training. The index comprises three risk groups (low, at zero to two risks; moderate, at three risks; and high, at four to five risks).

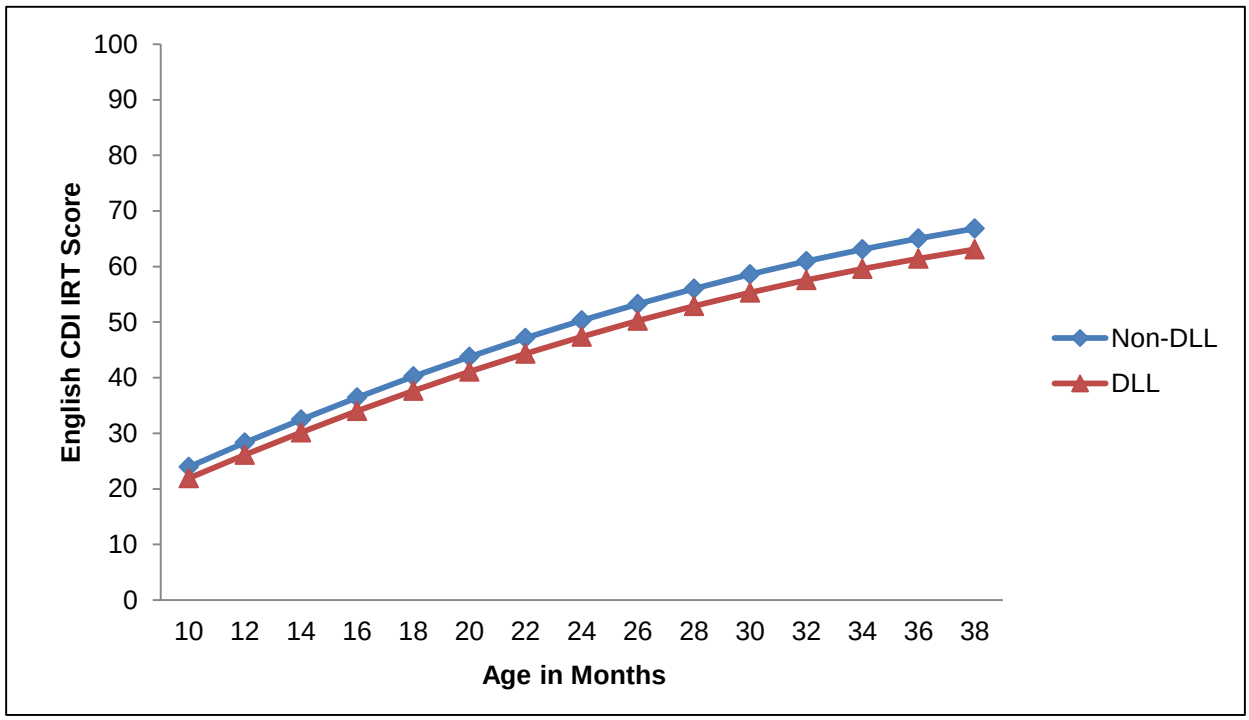
[b]Psychological risk is an index of cumulative risk based on (1) moderate or severe depressive symptoms, (2) reported parenting stress one standard deviation or higher than the sample mean on either the Parenting Stress subscale or the Parent-Child Dysfunctional Interaction subscale of the Parenting Stress Index, and (3) substance use problems including parent reports of drug use in the past year or having ever had a drug or drinking problem. Scores are classified as no risk (0 risk factors), (1 risk factor), and high (2 or 3 risk factors).

+$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$.

CDI= MacArthur-Bates Communicative Development Inventories, DLL=Dual language learner.

Very few child and family characteristics predicted children's average English vocabulary skills as measured by the CDI and none of the characteristics examined predict vocabulary growth. However, DLL status and maternal demographic risk were associated with average English vocabulary skills.[22] On average, DLL children score lower than children from English-speaking homes, but they grow at a rate similar to children from English-speaking homes (Figure IV.1). Children with medium or high maternal demographic risk score lower than children with lower risk yet risk is unrelated to growth.

Figure IV.1. Developmental Trajectories of Child English CDI by DLL Status



Source: Spring 2009 to spring 2012 Staff Child Report (SCR).

Note: Sample includes both the 1-year-old Cohort and Newborn Cohort children.

CDI= MacArthur-Bates Communicative Development Inventories.

There is linear growth[23] in Spanish-speaking children's vocabulary skills (Table IV.3). On average, boys score lower than girls. Children with low birth weight score marginally lower than children with normal birth weight (at trend level).[24] Counter-intuitively, Spanish-speaking children whose parents had two or more psychological risk factors grow at a faster rate than children whose parent had no psychological risk, which may be due in part to small sample sizes. On average, children in center-based programs score lower than children in multiple-approach programs, and grow at a marginally slower rate than children in multiple-approach programs (Figure IV.2). Children in programs offering only the home-based option, on average, score marginally higher than children in multiple-approach programs; children in home-based programs also grow at a marginally greater rate than children in multiple-approach programs.

[22] Because DLL status is confounded with race/ethnicity, we did not include race/ethnicity in the child-level model.

[23] The quadratic term is not significant and was dropped from the model.

[24] When we report marginal or trend level findings, this means the findings did not meet the threshold to be considered statistically significant ($p < .05$ or lower) but were close to or trending toward being significant ($p < .10$).

Table IV.3. Predictors of Children's Spanish Vocabulary Growth Trajectory

	Model 1	Model 2	Model 3
	Coefficient (Standard Error)	Coefficient (Standard Error)	Coefficient (Standard Error)
Mean Spanish CDI IRT Score	44.36 (1.30)***	47.99 (1.96)***	49.45 (2.02)***
Program service approach			
Center-based			-8.86 (1.95)***
Home-based			5.70 (3.29))+
Multiple-approach (referent)			
Male		-2.70 (1.30)*	-3.19 (1.32)*
Family income-to-needs ratio		-0.64 (0.89)	-0.58 (0.89)
Maternal demographic risk[a]			
Lower risk (referent)			
Moderate risk		-2.26 (1.94)	-2.58 (1.96)
High risk		0.96 (1.81)	0.61 (2.06)
Parent psychological risk[b]			
No risk (referent)			
One risk factor		1.04 (1.86)	0.84 (1.82)
Two or more risk factors		-0.37 (2.68)	-0.10 (2.66)
Low birth weight		-5.56 (2.34)*	-5.05 (2.76)+
Premature birth		-4.37 (3.43)	-5.17 (3.49)
Child with good or excellent health		1.32 (1.55)	1.72 (1.62)
Age Slope	1.26 (0.07)***	1.20 (0.14)***	1.28 (0.13)***
Program service approach			
Center-based			-0.40 (0.23)+
Home-based			0.21 (0.13)+
Multiple-approach (referent)			
Male		-0.05 (0.12)	-0.09 (0.11)
Family income-to-needs ratio		-0.04 (0.07)	-0.03 (0.06)
Maternal demographic risk			
Lower risk (referent)			
Moderate risk		0.17 (0.14)	0.15 (0.14)
High risk		-0.11 (0.20)	-0.12 (0.18)
Parent psychological risk			
No risk (referent)			
One risk factor		-0.10 (0.14)	-0.12 (0.12)
Two or more risk factors		0.73 (0.34)*	0.74 (0.35)*
Low birth weight		0.05 (0.30)	0.09 (0.25)
Pre term birth		0.47 (0.30)	0.44 (0.29)
Child with good or excellent health		0.05 (0.13)	0.02 (0.12)

Source: Spring 2009 to spring 2012 Staff Child Report (SCR) and Parent Interview.

Note: Sample includes both the 1-year-old Cohort and Newborn Cohort children.

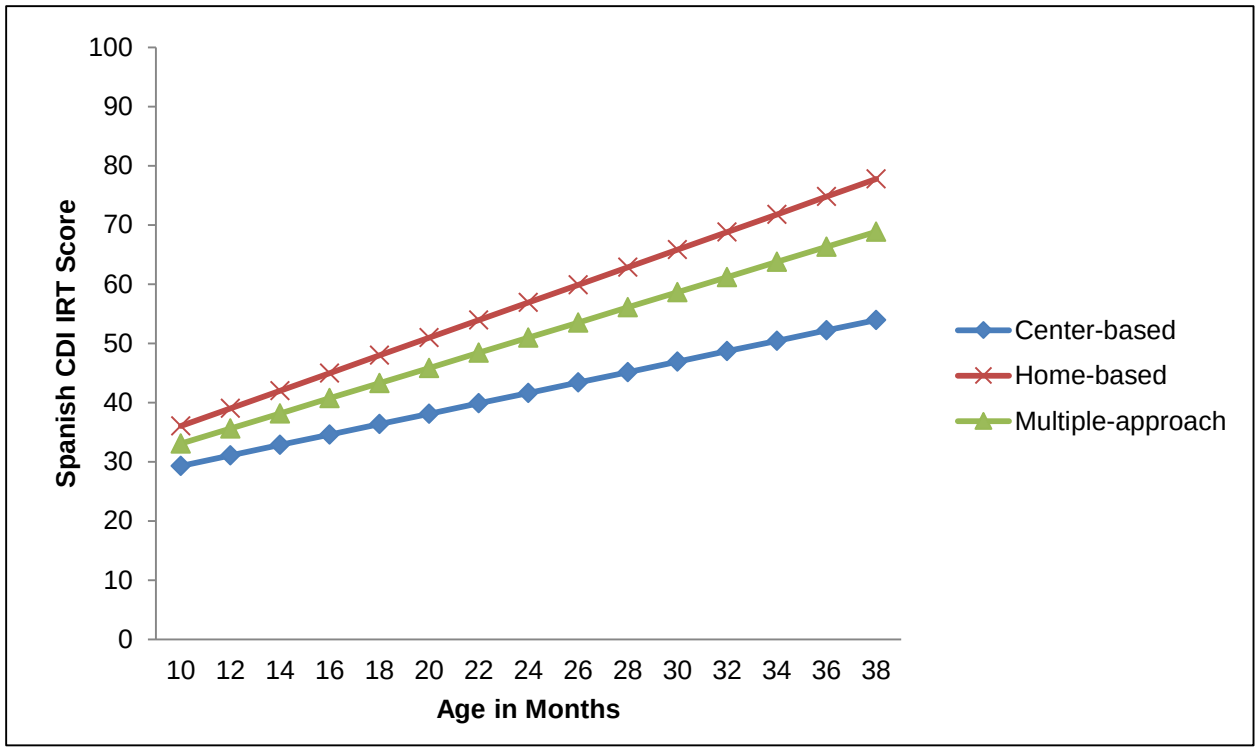
[a]Maternal demographic risk is the sum of five maternal characteristics collected at baseline: (1) single, (2) a teenager at first birth, (3) lacking a high school or equivalent credential, (4) receiving public assistance, or (5) not employed or in school or training. The index comprises three risk groups (low, at zero to two risks; moderate, at three risks; and high, at four to five risks).

[b]Psychological risk is an index of cumulative risk based on (1) moderate or severe depressive symptoms, (2) reported parenting stress one standard deviation or higher than the sample mean on either the Parenting Stress subscale or the Parent-Child Dysfunctional Interaction subscale of the Parenting Stress Index, and (3) substance use problems including parent reports of drug use in the past year or having ever had a drug or drinking problem. Scores are classified as no risk (0 risk factors), (1 risk factor), and high (2 or 3 risk factors).

+$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$.

CDI= MacArthur-Bates Communicative Development Inventories.

Figure IV.2. Developmental Trajectories of Child Spanish CDI by Program Approach



Source: Spring 2009 to spring 2012 Staff Child Report (SCR).

Note: Sample includes both the 1-year-old Cohort and Newborn Cohort children.

CDI= MacArthur-Bates Communicative Development Inventories.

Children's Social-Emotional Skills Improved over Early Head Start Enrollment

The BITSEA Competence scale assesses children's social-emotional abilities such as empathy, prosocial behaviors, and compliance. Higher scores indicate greater competence (see Box III.3 for more detailed descriptions of the measure). Results from growth curve analysis indicate that there is linear growth[25] in children's social competence during their time in Early Head Start according to parent reports (Table IV.4). On average, boys have lower social competence than girls; children with good or excellent health have greater social competence than children with poorer health; African American children are rated as having lower social competence than white children; children with high maternal demographic risk have marginally lower social competence than children with lower maternal risk (at a trend level); and children whose parents have one psychological risk factor have lower social competence than children whose parents have no psychological risk. None of the child/family characteristics predict growth rate in social competence.

With regard to program approach, children in programs offering only the center-based option have greater social competence than children in multiple-approach programs. The social competence of children in programs that offer only the home-based option increases at a faster rate than that of children in multiple-approach programs (Figure IV.3).

[25] The quadratic term is not significant and was dropped from the model.

Table IV.4. Predictors of Children's Social-Emotional Growth Trajectory

	Competence			Problem		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
	Coefficient (Standard Error)	Coefficient (Standard Error)	Coefficient (Standard Error)	Coefficient (Standard Error)	Coefficient (Standard Error)	Coefficient (Standard Error)
Mean Raw Score	16.91 (0.11)***	17.25 (0.26)***	17.18 (0.26)***	11.94 (0.32)***	10.20 (0.79)***	10.36 (0.79)**
Program service approach						
Center-based			0.57 (0.27)*			0.13 (0.57)
Home-based			0.17 (0.24)			-1.59 (1.13)
Multiple-approach (referent)						
Male		-0.88 (0.19)***	-0.86 (0.19)***		0.92 (0.46)*	0.94 (0.46)*
Race/ethnicity						
White (referent)						
Black		-0.77 (0.29)**	-0.82 (0.29)**		0.70 (0.62)	0.47 (0.61)
Hispanic		-0.19 (0.28)	-0.23 (0.28)		-0.29 (0.78)	-0.40 (0.77)
Other		-0.47 (0.39)	-0.47 (0.39)		0.20 (0.79)	0.01 (0.77)
DLL status		0.05 (0.27)	0.07 (0.27)		1.60 (0.74)*	1.54 (0.74)*
Family income-to-needs ratio		-0.01 (0.05)	-0.01 (0.05)		0.06 (0.14)	0.06 (0.14)
Maternal demographic risk[a]						
Lower risk (referent)						
Moderate risk		-0.12 (0.22)	-0.15 (0.22)		1.00 (0.44)*	1.03 (0.43)*
High risk		-0.51 (0.27)+	-0.53 (0.26)*		1.72 (0.58)**	1.76 (0.58)**
Parent psychological risk[b]						
No risk (referent)						
One risk factor		-0.68 (0.28)*	-0.69 (0.27)*		2.50 (0.54)***	2.47 (0.56)***
Two or more risk factors		-0.40 (0.38)	-0.40 (0.38)		5.47 (0.94)***	5.41 (0.93)***
Low birth weight		-0.41 (0.45)	-0.38 (0.45)		1.13 (1.31)	1.17 (1.31)
Premature birth		-0.58 (0.50)	-0.57 (0.50)		0.11 (1.06)	0.15 (1.06)
Child with good or excellent health		1.06 (0.24)***	1.06 (0.24)***		-1.73 (0.59)**	-1.65 (0.59)**
Age Slope	0.07 (0.01)***	0.10 (0.02)***	0.10 (0.02)***	0.04 (0.01)**	0.03 (0.04)	0.04 (0.04)
Program service approach						
Center-based			-0.02 (0.04)			-0.07 (0.03)*
Home-based			0.05 (0.02)**			-0.00 (0.04)
Multiple-approach (referent)						
Male		0.00 (0.02)	0.00 (0.02)		0.02 (0.02)	0.02 (0.02)
Race/ethnicity						
White (referent)						
Black		-0.03 (0.02)	-0.02 (0.02)		-0.03 (0.04)	-0.02 (0.05)
Hispanic		-0.00 (0.02)	0.00 (0.02)		0.02 (0.05)	0.03 (0.05)
Other		-0.00 (0.02)	0.01 (0.02)		0.02 (0.06)	0.02 (0.06)
DLL status		-0.03 (0.02)	-0.03 (0.02)		-0.03 (0.05)	-0.03 (0.05)
Family income-to-needs ratio		-0.00 (0.00)	-0.00 (0.00)		0.01 (0.01)	0.01 (0.01)
Maternal demographic risk[a]						
Lower risk (referent)						
Moderate risk		0.00 (0.02)	0.00 (0.02)		0.00 (0.03)	0.00 (0.03)

	Competence			Problem		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
	Coefficient (Standard Error)	Coefficient (Standard Error)	Coefficient (Standard Error)	Coefficient (Standard Error)	Coefficient (Standard Error)	Coefficient (Standard Error)
High risk		-0.03 (0.02)	-0.03 (0.02)		0.08 (0.04)+	0.09 (0.04)+
Parent psychological risk[b]						
No risk (referent)						
One risk factor		-0.01 (0.02)	-0.01 (0.02)		0.01 (0.03)	0.00 (0.03)
Two or more risk factors		0.01 (0.03)	0.01 (0.03)		0.02 (0.07)	0.02 (0.07)
Low birth weight		0.02 (0.04)	0.02 (0.04)		0.02 (0.06)	0.01 (0.06)
Premature birth		-0.02 (0.03)	-0.02 (0.03)		0.02 (0.06)	0.02 (0.06)
Child with good or excellent health		-0.02 (0.02)	-0.02 (0.02)		-0.02 (0.04)	-0.02 (0.04)
Squared Age	--	--	--	-.008 (.002)***	-.008 (.002)***	-.008 (.002)***

Source: Spring 2009 to spring 2012 Parent Interview.

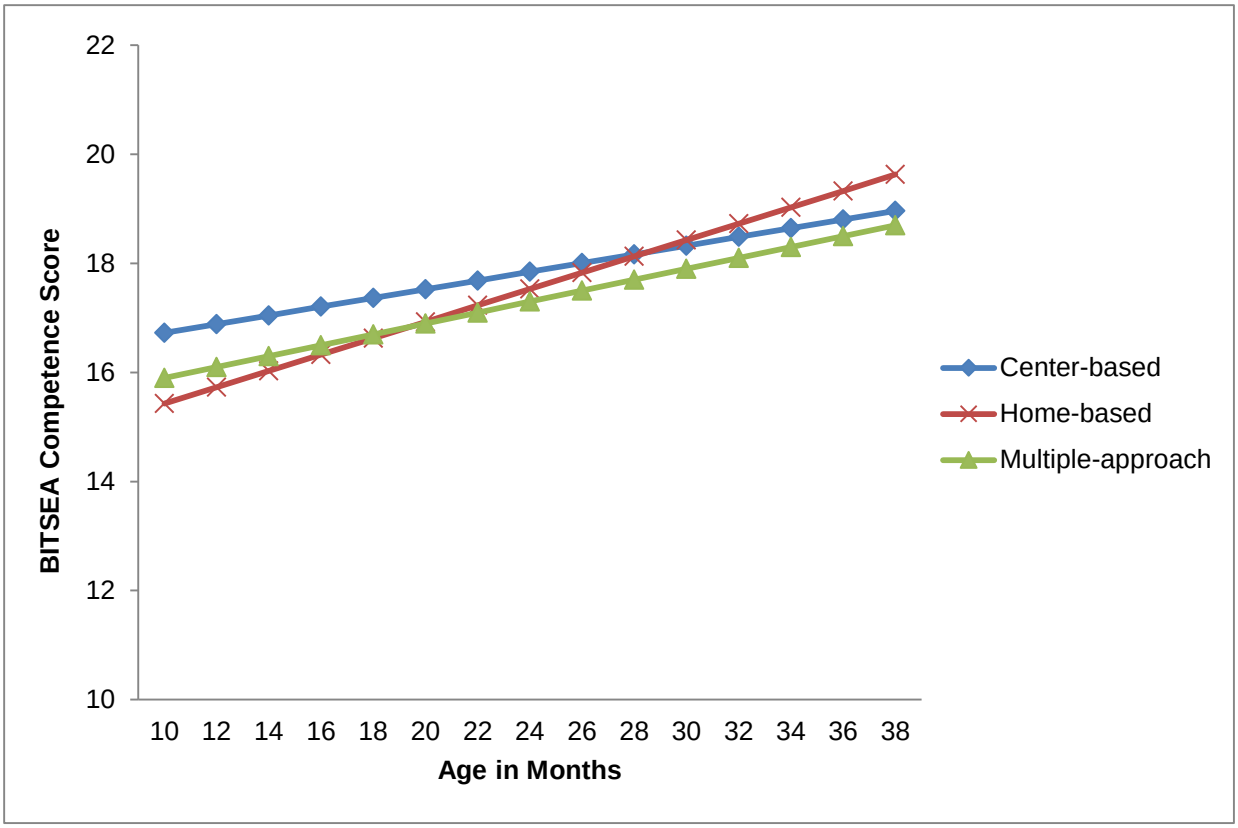
Note: Sample includes both the 1-year-old Cohort and Newborn Cohort children. Children's social-emotional development was assessed using the Brief Infant Toddler Social Emotional Assessment (BITSEA).

[a]Maternal demographic risk is the sum of five maternal characteristics collected at baseline: (1) single, (2) a teenager at first birth, (3) lacking a high school or equivalent credential, (4) receiving public assistance, or (5) not employed or in school or training. The index comprises three risk groups (low, at zero to two risks; moderate, at three risks; and high, at four to five risks).

[b]Psychological risk is an index of cumulative risk based on (1) moderate or severe depressive symptoms, (2) reported parenting stress one standard deviation or higher than the sample mean on either the Parenting Stress subscale or the Parent-Child Dysfunctional Interaction subscale of the Parenting Stress Index, and (3) substance use problems including parent reports of drug use in the past year or having ever had a drug or drinking problem. Scores are classified as no risk (0 risk factors), (1 risk factor), and high (2 or 3 risk factors). +$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$.

DLL=Dual language learner

Figure IV.3. Developmental Trajectories of Child Social Competence by Program Approach

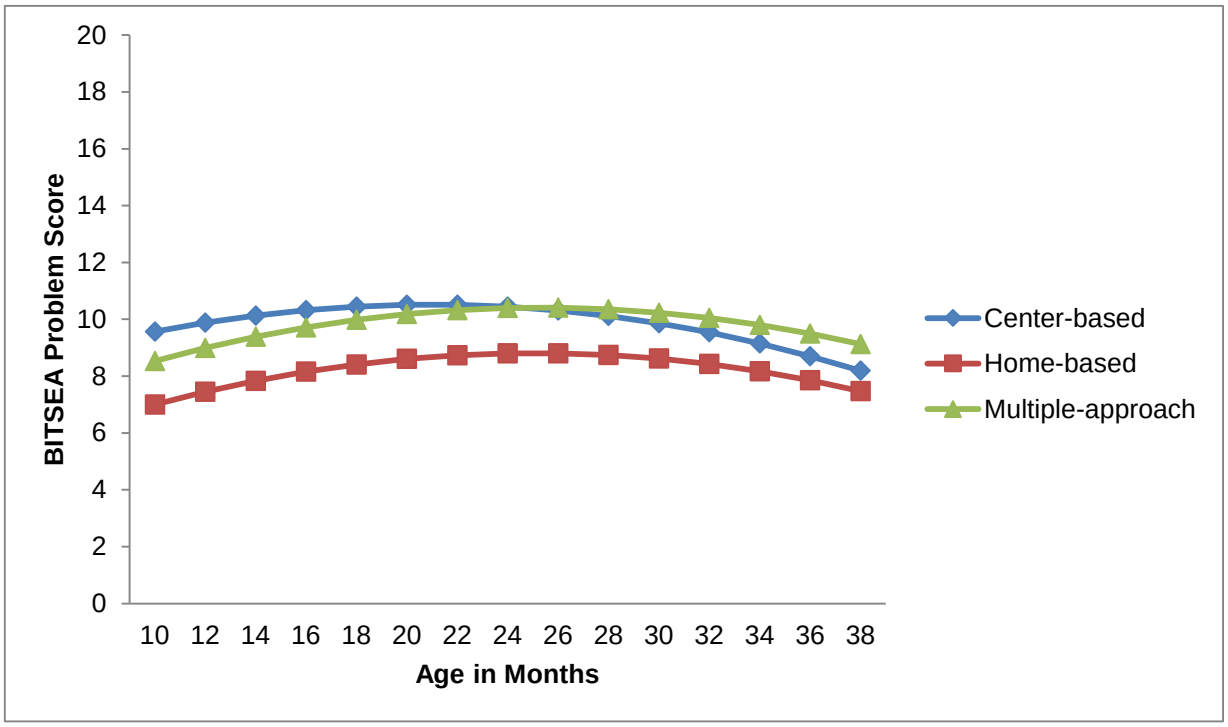


Source: Spring 2009 to spring 2012 Parent Interview.

Note: Sample includes both the 1-year-old Cohort and Newborn Cohort children.

The BITSEA Problem scale assesses social-emotional and behavioral problems such as aggression, defiance, overactivity, negative emotionality, anxiety, and withdrawal. Higher scores indicate more problems. (See Box III.3 for more detailed descriptions of the measure.) Growth curve analysis results suggest that the change in children's problem behaviors during the Early Head Start years is non-linear (Table IV.4) Children's problem behaviors increase over time but the rate of increase appears to slow down over time. On average, boys have more problem behaviors than girls; DLL children were rated as having more problem behaviors than non-DLL children; parents with a greater number of maternal demographic and psychological risk factors tend to rate their children as having more problem behaviors; and children with good or excellent health are rated as having fewer problem behaviors than children with poorer health. The trajectories of change in problem behaviors are different for children in programs offering different program approaches (Figure IV.4). Problem behaviors decrease over time for children in programs offering only the center-based option, and remain the same for children in multiple-approach programs.

Figure IV.4. Developmental Trajectories of Child Problem Behaviors by Program Approach



Source: Spring 2009 to spring 2012 Parent Interview.

Note: Sample includes both the 1-year-old Cohort and Newborn Cohort children.

Parent Mental Health and Parenting Stress Improved over Early Head Start Enrollment

The parent outcome measures in Baby FACES are available at four time points for the Newborn Cohort (age 0 to age 3), at three time points for the 1-year-old Cohort (age 1 to age 3) for the Center for Epidemiologic Studies Depression Scale-Short Form (CESD-SF), and three time points (age 1 to age 3) for the Parenting Stress Index–Short Form (PSI-SF). The CESD-SF assesses parent symptoms of depression or psychological distress. The PSI-SF measures the degree of stress in parent-child relationships using two subscales: Parental Distress and Parent-Child Dysfunctional Interaction. (See Chapter II for descriptions of the CESD-SF and PSI-SF.) In growth curve modeling of these parent outcomes, we did not include parent psychological risk factors as covariates because they are components of the risk indices. We expect that parent depressive symptoms, parental distress, and parent-child dysfunctional interaction will decrease over time in Early Head Start. The bottom section of Table IV.1 shows the descriptive statistics for the CESD-SF and PSI-SF subscale scores at each point of data collection. There is an overall trend of decrease over time in both parent depressive symptoms and parenting stress.

Parent Depressive Symptoms Diminished

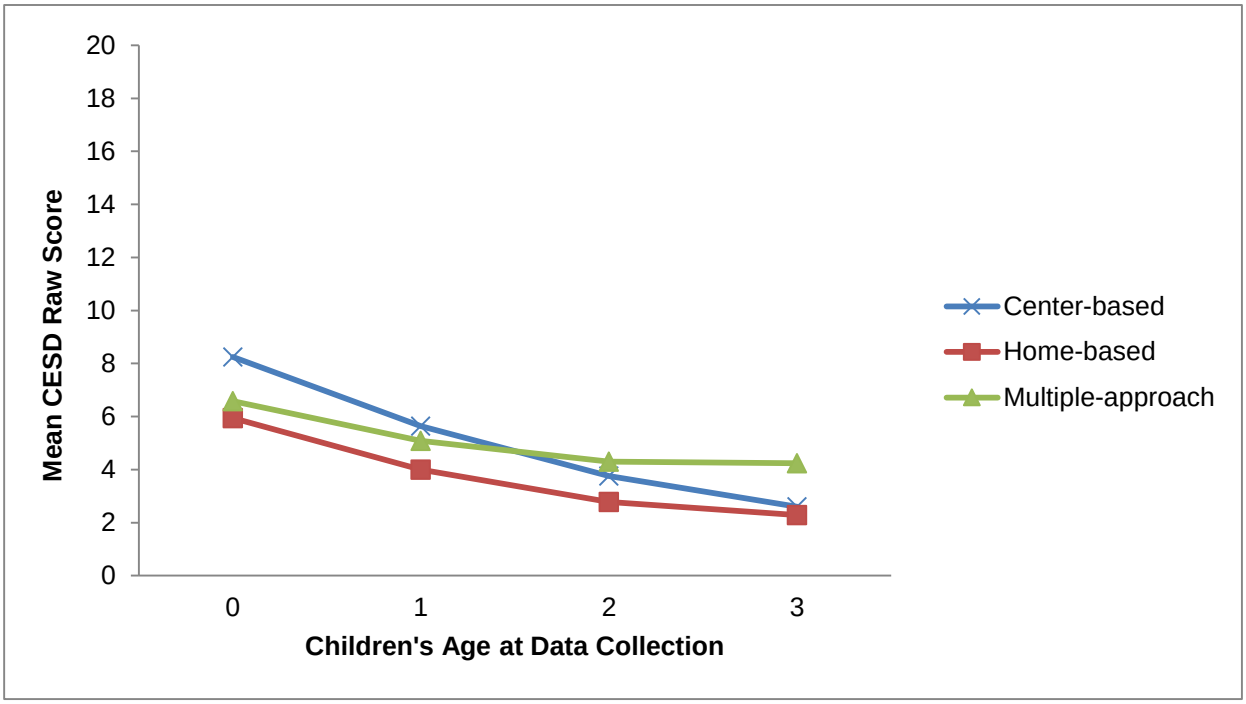
Growth curve analysis shows that parent depressive symptoms decrease over time in a non-linear fashion (Table IV.5). The rate of decrease slows down over time. Hispanic parents report fewer depressive symptoms than white parents on average. Parents' depressive symptoms decreased at a faster rate in center-based programs than in multiple-approach programs (Figure IV.5).

Levels of Parenting Stress Decreased

Growth curve analysis shows that there is a linear decrease in Parental Distress during the time families were enrolled in Early Head Start (Table IV.5). On average, parents of Hispanic or DLL children and parents from poorer families reported higher parental distress. Hispanic parents' levels of parental distress decrease at a marginally greater rate than white parents (at a trend level). With regard to program approach, parents in multiple-approach programs reported higher levels of parental distress than parents in other programs.

On average, there is no significant change in Parent-Child Dysfunctional Interaction; however, the average levels of Parent-Child Dysfunctional Interaction differ by child and family characteristics. Parents of African American, Hispanic, or DLL children and parents with high maternal risk reported higher levels of parent-child dysfunctional interaction on average than those in the counterpart groups. Parents in center-based programs reported marginally lower levels of parent-child dysfunctional interaction than parents in multiple-approach programs (at a trend level).

Figure IV.5. Change Trajectories of Parent Depressive Symptoms by Program Approach



Source: Spring 2009 to spring 2012 Parent Interview.

Note: Sample includes parents of both the 1-year-old Cohort and Newborn Cohort children.

Table IV.5. Predictors of Change Trajectory for Parent Outcomes

	PSI: Parental Distress[a]			PSI: Parent-Child Dysfunctional Interaction[a]			CESD: Depressive Symptoms[b]		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)
Mean Raw Score	10.68 (0.23)***	9.63 (0.36)***	9.88 (0.37)***	8.47 (0.21)***	7.25 (0.28)***	7.31 (0.30)***	5.17 (0.23)***	4.98 (0.40)***	5.08 (0.41)***
Program service approach									
Center-based			-0.86 (0.32)**			-0.52 (0.32)+			0.56 (0.67)
Home-based			-1.34 (0.29)***			-0.17 (0.21)			-1.08 (0.72)
Multiple-approach (referent)									
Male		0.24 (0.32)	0.24 (0.32)		-0.23 (0.32)	-0.24 (0.32)		0.48 (0.34)	0.51 (0.34)
Race/ethnicity									
White (referent)									
Black		0.64 (0.42)	0.58 (0.41)		0.94 (0.40)*	0.99 (0.40)*		-0.46 (0.58)	-0.68 (0.58)
Hispanic		1.03 (0.56)+	1.03 (0.56)+		0.96 (0.44)*	1.01 (0.44)*		-1.28 (0.59)*	-1.39 (0.59)*
Other		0.41 (0.74)	0.33 (0.73)		0.99 (0.51)+	1.00 (0.52)+		0.48 (0.76)	0.38 (0.75)
DLL status		1.26 (0.42)**	1.19 (0.42)**		1.23 (0.38)**	1.22 (0.38)**		0.20 (0.56)	0.19 (0.56)
Family income-to-needs ratio		-0.08 (0.03)*	-0.08 (0.03)*		-0.02 (0.03)	-0.02 (0.03)		-0.06 (0.05)	-0.06 (0.05)
Maternal demographic risk[a]									
Lower risk (referent)									
Moderate risk		-0.24 (0.41)	-0.19 (0.41)		0.25 (0.36)	0.26 (0.36)		0.70 (0.51)	0.73 (0.51)
High risk		0.06 (0.53)	0.08 (0.53)		1.03 (0.52)*	1.03 (0.51)*		0.70 (0.58)	0.73 (0.57)
Time Slope	-0.49 (0.17)**	-0.39 (0.34)	-0.38 (0.34)	0.02 (0.12)	0.10 (0.24)	0.10 (0.24)	-1.26 (0.32)**	-1.33 (0.44)**	-1.14 (0.44)*
Program service approach									
Center-based			--			--			-1.10 (0.39)**
Home-based			--			--			-0.44 (0.39)
Multiple-approach (referent)									
Male		0.42 (0.27)	0.40 (0.27)		0.22 (0.26)	0.21 (0.26)		-0.16 (0.23)	-0.21 (0.24)

	PSI: Parental Distress[a]			PSI: Parent-Child Dysfunctional Interaction[a]			CESD: Depressive Symptoms[b]		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Race/ethnicity									
White (referent)									
Black		-0.12	-0.12		-0.12	-0.11		0.66	0.75
		(0.38)	(0.38)		(0.30)	(0.30)		(0.62)	(0.62)
Hispanic		-0.69	-0.70		-0.33	-0.33		-0.13	-0.08
		(0.39)+	(0.39)+		(0.30)	(0.30)		(0.39)	(0.39)
Other		0.13	0.11		-0.05	-0.06		0.15	0.11
		(0.63)	(0.63)		(0.60)	(0.60)		(0.49)	(0.50)
DLL status		-0.21	-0.21		-0.03	-0.03		-0.18	-0.23
		(0.33)	(0.33)		(0.27)	(0.27)		(0.34)	(0.35)
Family income-to-needs ratio		0.00	0.00		-0.00	-0.00		0.03	0.02
		(0.06)	(0.06)		(0.05)	(0.05)		(0.06)	(0.06)
Maternal demographic risk[c]									
Lower risk (referent)									
Moderate risk		-0.18	-0.17		-0.28	-0.28		-0.04	-0.02
		(0.37)	(0.37)		(0.29)	(0.28)		(0.35)	(0.35)
High risk		0.19	0.21		0.17	0.17		0.58	0.59
		(0.40)	(0.40)		(0.38)	(0.38)		(0.44)	(0.44)
Squared Time	--	--	--	--	--	--	0.35	0.37	0.36
							(0.19)+	(0.19)*	(0.19)+

Source: Spring 2009 to spring 2012 Parent Interview.

Note: Sample includes parents of both the 1-year-old Cohort and Newborn Cohort children. Parents' depressive symptoms were assessed using the Center for Epidemiologic Studies Depression Scale-Short Form (CESD-SF). Parenting stress was assessed using the Parenting Stress Index–Short Form (PSI-SF), which includes two subscales: Parental Distress and Parent-Child Dysfunctional Interaction.

[a]Models for the two PSI subscales include 3 time points.
[b]Models for parents' depressive symptoms include 4 time points.
[c]Maternal demographic risk is the sum of five maternal characteristics collected at baseline: (1) single, (2) a teenager at first birth, (3) lacking a high school or equivalent credential, (4) receiving public assistance, or (5) not employed or in school or training. The index comprises three risk groups (low, at zero to two risks; moderate, at three risks; and high, at four to five risks).

+$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$.

SE = Standard error; DLL=Dual language learner.

Summary of Key Findings

- Children show growth in language and social-emotional domains.
  - Children's receptive and expressive vocabulary skills increase over the period of enrollment in Early Head Start.
    - On average, DLL children score lower on English vocabulary than children from English-speaking homes, but they grow at a rate similar to children from English-speaking homes; children with high maternal demographic risk score lower than children with lower risk.
    - On average, boys score lower than girls on Spanish vocabulary and children in center-based programs score lower than children in home-based and multiple-approach programs.
  - Children's social-emotional skills improve over the period of enrollment in Early Head Start.
    - Child's gender, race/ethnicity, general health, maternal demographic risk and psychological risk are associated with children's social competence. Program approach is associated with the average level as well as the growth rate of social competence.
    - Child's gender, DLL status, general health, and maternal demographic and psychological risks are associated with children's problem behaviors. Problem behaviors decrease over time for children in programs offering only the center-based option, and remain the same for children in multiple-approach programs.
- Parent mental health and the levels of parenting stress improve over the period of enrollment in Early Head Start.
  - Both parent depressive symptoms and parenting stress decline over the period of enrollment in Early Head Start.
    - Hispanic parents reported fewer depressive symptoms than white parents on average. Parents' depressive symptoms decrease at a faster rate in center-based programs than in multiple-approach programs.
    - Parents of Hispanic or DLL children and parents from poorer families reported higher parental distress. Parents in multiple-approach programs reported higher levels of parental distress than parents in other programs.
    - On average, there is no significant change in Parent-Child Dysfunctional Interaction; however, the average levels of parent-child dysfunctional interaction differ by race/ethnicity, DLL status, and maternal demographic risk.

V. PROGRAM AND STAFF CHARACTERISTICS, QUALITY OF SERVICES OFFERED, AND PROGRAM IMPLEMENTATION

An important aspect of the Early Head Start conceptual framework (see Figure I.1) is the relationship between the provision of high-quality services and children's growth and development. The quality of early childhood services is multidimensional and encompasses not only characteristics of staff and the physical environment but also the quality of the interactions and relationships among staff members and the children and parents with whom they work. Accordingly, Baby FACES provides important, multifaceted information about the overall quality of classrooms and home-based services including the characteristics of teachers and home visitors serving children and families, characteristics of the physical classroom environment (that is, structural characteristics), and the quality of the interactions and relationships among staff members and the children and parents with whom they work (that is, process characteristics).

This chapter uses data provided by teachers, home visitors, and program directors to describe characteristics of Early Head Start staff serving study children.[26] Program directors reported on staffing characteristics of the program they oversee as a whole, but teaching and home visiting staff also provided detailed information about themselves.[27] Here, we provide reports from all three respondents–program directors, teachers, and home visitors–on Early Head Start program staff characteristics. Next, we provide information on key aspects of children's home visits, and we describe both structural and process characteristics of children's classrooms. We begin this chapter by focusing on program director reports on staffing. This descriptive information on Early Head Start staff and home visit and classroom quality serve as the foundation for the hierarchical linear modeling discussed in Chapter VI. Because we did not sample classrooms or home visitors, we cannot aggregate quality measures to the program level and cannot make inferences about what is typical quality for a given center or program. We can only describe the experiences of sample children at age 3 across programs. See Chapter II and Appendix A for details on sampling procedures.

Staffing, Retention, and Training Contribute to Quality of Early Head Start Programs

To examine teacher and home visitor characteristics that may influence the quality of child care, the teacher and home visitor interviews provide background information including socio-demographic details, education and training, and professional experience. Program directors provided staffing-related information on the program as a whole, including the number of teachers and home visiting staff; efforts to improve quality through professional development; and issues associated with the retention of frontline and management staff.[28] This chapter details the information reported by program directors on overall

[26] Teacher and home visitor data are from spring 2011 for children at age 3 in the 1-year-old Cohort and from spring 2012 for children at age 3 in the Newborn Cohort. Data from program directors are from spring 2011 and include programs serving children in both the 1-year-old and Newborn Cohorts.

[27] Throughout this chapter, we distinguish between data from this smaller group of individual staff members and aggregate data about the program as reported by program directors.

[28] Early Head Start programs employ a variety of frontline and management staff members. These include, but are not limited to: teachers, home visitors, directors, assistant directors, managers, coordinators, and specialists. Teachers include all staff with primary responsibility for all or some of the children in a classroom, and home visitors include all staff whose primary function is to make regular home visits to families and children. Frontline staff members include all staff who work directly with children and families, which typically includes teachers in center-based programs, home visitors in home-based programs, and both in multiple-approach programs. Management staff are responsible for monitoring programs' progress toward goals and overseeing implementation of program services.

staffing, retention, and training, and then discusses the characteristics of teachers and home visitors serving children in Baby FACES.

Programs Have Moderately Low Staff Turnover Rates

Program directors reported that 15 percent of teachers and 10 percent of home visitors left the program in the past year (see Table V.1).[29] This is similar to the previous year when these rates were 15 and 10 percent, respectively.[30] Nearly two-thirds of directors reported that these teachers and home visitors left for personal reasons (63 percent). Nearly half of the directors (45 percent) reported teachers and home visitors left for a change in career. About one-third of directors reported the staff left due to a firing or layoff (37 percent) or for higher compensation or improved benefits (34 percent).[31] Turnover at the leadership level in programs was higher but not as high as the previous year, with 28 percent of programs losing a coordinator or manager, and 15 percent losing a director (in the previous year 43 percent lost a coordinator or manager and 17 percent lost a director). On average, each program lost about 1 member of the leadership staff during the past year.

Table V.1. Staff Turnover Is Moderately Low

	Weighted Percentage or Mean (Standard Error)
Mean Number of Teachers Currently Employed	22.9 (3.10)
Mean Number of Home Visitors Currently Employed	7.7 (0.67)
Turnover Rate Among Frontline Staff (percentage of staff who left program in past 12 months)	
Teachers	14.6 (2.66)
Home visitors	10.2 (1.57)
Reasons Teachers and/or Home Visitors Left (percentage of programs)[a]	
Personal reasons	63.2 (6.41)
Change in careers	44.7 (7.01)
Firing/layoff	36.7 (6.76)
Higher compensation/better benefits	34.4 (6.72)
Maternity leave	11.6 (3.55)
Other	25.5 (6.04)
Percentage of Programs in Which a Coordinator or Manager Left in the Past 12 Months	28.3 (5.41)
Percentage of Programs in Which a Director Left in the Past 12 Months	14.6 (4.25)
Mean Number of Coordinators or Managers Who Left in the Past 12 Months[a]	0.5 (0.12)
Sample Size	**76-89**

Source: Spring 2011 Program Director Interview.

[a]Among programs in which teachers/home visitors, coordinators, or managers left. Reasons reported for staff departures from the program are not mutually exclusive.

[29] Directors reported that turnover rates among teachers ranged from 0 to 88 percent, while among home visitors it ranged from 0 to 50 percent. Thirty-five percent of programs had no teacher turnover, and 58 percent had no turnover of home visitors.

[30] These rates of turnover are low compared with other studies of frontline staff. In the EHSREP in fall 1999, turnover in most Early Head Start research programs ranged between 15 and 32 percent. It was 39 percent in the 11 programs that offered all or some center-based care (ACF 2002b).

[31] Reasons reported for staff departures from the program are not mutually exclusive.

One-third (33 percent) of programs have unfilled full-time staff positions (Table V.2). This percentage is lower than in spring 2010, when almost half had vacant positions, but similar to percentages in spring 2009. On average, program directors report that they have about two unfilled full-time staff positions. Most commonly, programs have vacancies for teachers (54 percent), home visitors (27 percent), and managers/supervisors (16 percent). Seven percent of programs have unfilled director positions, which is about twice the percentage of director vacancies as in spring 2009 but similar to percentages in spring 2010 (8 percent).

Table V.2. One-Third of Programs Have Unfilled Full-Time Staff Positions

	Weighted Percentage or Mean (Standard Error)
Program Currently Has at Least One Unfilled Full-Time Staff Position (percentage of programs)	32.9 (5.48)
Mean Number of Full-Time Positions Currently Unfilled[a]	2.3 (0.24)
Type of Position Unfilled (percentage of programs)[a]	
Teacher	53.5 (8.78)
Home visitor	26.9 (6.58)
Manager/supervisor	15.9 (7.28)
Director	7.2 (3.85)
Other	35.7 (7.90)
Level of Staff Salaries and Benefits (percentage of programs)	
Below the average in the surrounding area	22.7 (5.57)
About the same as the average in the surrounding area	42.0 (6.39)
Above the average in the surrounding area	35.3 (5.56)
Sample Size	**89**

Source: Spring 2011 Program Director Interview.

[a]Among programs with unfilled full-time positions.

Despite staff-retention problems, programs generally report that salaries and benefits are commensurate with other early childhood education positions in the surrounding area (Table V.2). In fact, less than one-quarter of programs (23 percent) pay salaries and benefits below the average for the surrounding area, according to program directors. Forty-two percent of programs pay salaries and benefits that are about the same as the surrounding area, and 35 percent pay above the average. This finding marks a shift from patterns in spring 2010, when a higher proportion of programs offered above-average salaries and benefits.

Staff Participate in a Number of Training and Professional Development Activities

As reported by program directors, nearly all programs (99 percent) develop staff training plans each year, with all programs soliciting feedback from staff on their needs, specifically for training new staff members (Table V.3). All programs also offer specialized training for new staff members. Program directors reported that nearly all frontline staff (98 percent) attended at least three training sessions in the past year. In one program, the director reported that only 50 percent of frontline staff completed at least three training sessions in the past year and in six programs the percent of staff completing at least three trainings was between 80 and 98. A later section of this chapter describes the participation of teachers and home visitors in these professional development activities.

Table V.3. Staff Training Opportunities Are Widely Available

	Weighted Percentage or Mean (Standard Error)
Program Develops Staff Training Plan Each Year (percentage of programs)	98.8 (1.20)
Program Solicits Information on Staff Needs to Inform Training Plan (percentage of programs)	100.0 (0.00)
Mean Percentage of Frontline Staff in Programs Who Attended at Least Three Training Sessions in Past Year	98.2 (1.03)
Program Offers Specialized Training for New Staff Members (percentage of programs)	99.6 (0.40)
Sample Size	**89**

Source: Spring 2011 Program Director Interview.

Background Characteristics of Teachers and Home Visitors

Research has linked some aspects of staff characteristics to child outcomes, and indeed, Early Head Start's conceptual framework illustrates such a pathway (Figure I.1). Important staff characteristics associated with the quality of care include amount and type of education and training, beliefs, and job satisfaction (Burchinal et al. 2000). For example, teacher education has been associated with children's cognitive and social-emotional development (Burchinal et al. 1996; Clarke-Stewart 1989; Hayes et al. 1990; Ruopp et al. 1979; Whitebook et al. 1989; Zaslow 1991), although these relations are typically weak, especially in more recent studies (Bogard et al. 2008; Early et al. 2006). In this section, we use teacher and home visitor reports to describe the characteristics of teachers and home visitors working with study children.[32]

Children Have Diverse Teachers and Home Visitors

Nearly all study children have a teacher (99 percent) or home visitor (99 percent) who is female (see Table V.4). About half (49 percent) of those in center-based settings have a teacher who is white, whereas 61 percent of those in the home-based option have a home visitor who is white. About one-quarter of children in center-based settings have a teacher who is Hispanic (27 percent), and another 17 percent have an African American teacher. Seven percent have a teacher from other racial/ethnic backgrounds. Among those receiving home-based services, more than one-quarter have a Hispanic home visitor (29 percent), 7 percent have an African American home visitor, and 3 percent have a home visitor from another racial/ethnic background.

[32] Because we did not sample teachers or home visitors, we must provide overall estimates as a percentage of children having staff with specific attributes, rather than as a percentage of teachers and home visitors. Chapter II and Appendix A provide information on the sample and sampling procedures.

Table V.4. Children Have Diverse Teachers and Home Visitors

Characteristics	Weighted Mean or Percentage (Standard Error)	
	Teachers	Home Visitors
Female (percentage)	99.4 (0.64)	98.7 (0.99)
Race/Ethnicity (percentage)		
White, non-Hispanic	49.3 (6.14)	61.4 (6.40)
African American, non-Hispanic	17.3 (4.58)	7.3 (3.53)
Hispanic/Latino	26.8 (6.22)	28.7 (6.08)
Other, non-Hispanic	6.6 (2.50)	2.6 (1.40)
Sample Size	**229-313**	**195-213**

Sources: Spring 2011 and 2012 Teacher and Home Visitor Interviews.

Note: Includes data on Early Head Start staff serving children in both cohorts. Because we did not sample teachers or home visitors, we must provide overall estimates as a percentage of children rather than as a percentage of teachers and home visitors.

English and Spanish Are the Languages most Commonly Spoken in Classrooms and During Home Visits

Many children have a teacher or home visitor who speaks a language in addition to English, and for both types of staff, that language is likely to be Spanish (see Table V.5). Overall, 34 percent of children receiving home-based services and 31 percent of those receiving center-based services have a home visitor or teacher speaking a language in addition to English.[33] Similar percentages of children receiving home-based services (34 percent) and center-based services (31 percent) have a Spanish-speaking home visitor or teacher. Nine percent of children have a home visitor and 5 percent have a classroom teacher who speaks a language other than English or Spanish.

Similar to spring 2009 and 2010, children served in the center-based option are in classrooms in which three children, on average, speak a language other than English. Meanwhile, children have home visitors who reported about three families in their caseload speak a language other than English.[34] The language backgrounds of these children are similar to what we saw in spring 2009 and 2010. For example, in classrooms and home-based services, children most frequently speak Spanish (86 and 84 percent, respectively). In classrooms, Asian languages are the next most commonly spoken (5 percent); in home-based options, the next language most commonly spoken by families is Arabic (11 percent).[35]

[33] This percentage is an average across all children, not conditioned on those who speak a language other than English.

[34] Teachers report an average of 7.7 children currently enrolled in their classrooms. Home visitors report an average of 9.5 families in their caseloads.

[35] Between 11 and 14 percent (home-based and center-based, respectively) report that families speak approximately 10 languages other than those listed by name.

Table V.5. Many Children, Families, and Staff Speak a Language Other than English

Characteristics	Weighted Mean or Percentage (Standard Error)	
	Teachers	Home Visitors
Speaks Language Other than English (percentage)	31.2 (5.65)	34.1 (6.39)
Spanish	31.5 (5.55)	34.0 (6.17)
Other	5.1 (2.43)	8.6 (2.81)
Mean Number of Families per Classroom/Caseload Speaking		
English only	5.4 (0.43)	6.1 (0.58)
English and another language	1.8 (0.32)	1.6 (0.24)
Only another language	0.7 (0.46)	1.8 (0.47)
Non-English Languages Spoken by Families (percentage)		
Spanish	86.1 (5.93)	83.8 (6.02)
Arabic	2.2 (1.36)	11.3 (4.84)
Asian languages	5.4 (2.81)	6.3 (4.82)
Other	15.1 (3.89)	27.7 (7.25)
Languages Used for Communication During Home Visits (percentage)		
English	n.a.	90.7 (4.21)
Spanish	n.a.	45.6 (7.20)
Other	n.a.	11.2 (4.84)
Languages Spoken by Adults[a] in Classroom (percentage)		
English	100.0 (0.00)	n.a.
Spanish	47.2 (6.00)	n.a.
Other	13.6 (3.81)	n.a.
Non-English Language Spoken in Classroom by (percentage)		
Teacher	44.2 (5.35)	n.a.
Assistant teacher	24.3 (4.62)	n.a.
Classroom aide	0.0 (0.00)	n.a.
Volunteer/Nonstaff	0.0 (0.00)	n.a.
Language Used Most Often to Read to Children in Classroom (percentage)		
English	95.7 (2.91)	n.a.
Spanish	4.3 (2.91)	n.a.
Other	0.0 (0.00)	n.a.
Teacher/Home Visitor Communicates with Families Speaking Non-English Languages (percentage)		
Communicates only in English	61.3 (4.61)	58.9 (6.41)
Uses an informal interpreter	80.8 (6.09)	64.3 (6.06)
Uses physical cues or hand gestures	81.2 (3.85)	78.2 (4.45)
Uses bilingual newsletters/flyers/handouts	84.7 (3.75)	80.2 (4.87)
Uses pictures/drawing pictures	49.4 (5.31)	48.5 (5.64)
Uses books/dictionary	70.1 (5.36)	58.3 (5.93)
Uses other methods	14.0 (3.23)	16.0 (3.74)
Sample Size	**177-313**	**146-213**

Sources: Spring 2011 and 2012 Teacher and Home Visitor Interviews.

Note: Includes data on Early Head Start staff serving children in both cohorts. Because we did not sample teachers or home visitors, we must provide overall estimates as a percentage of children rather than as a percentage of teachers and home visitors.

[a]Adults include the lead teacher, assistant teacher, classroom aide, or volunteer/non-staff. Language categories do not sum to 100 because more than one language could be spoken in a classroom.

n.a. = not applicable.

English is the language most often spoken by adults[36] in classrooms and during home visits. All 3-year-olds in the center-based option have teachers who reported that adults use English in the classroom. Some children are exposed to more than one language, for instance, 47 percent of center-based children are in classrooms in which Spanish is also reported to be spoken by adults. Among center-based staff, teaching staff are the ones most likely to use a language in addition to or other than English in the classroom (44 percent of teachers and 24 percent of assistant teachers). Classroom aides and volunteer/non-staff members are not reported as speaking languages other than English. Nearly all children (96 percent) are in classrooms in which English is most often used when reading to children. Spanish is most often used for the other 4 percent.

For children in the home-based option, 91 percent have home visitors who use English during home visits. Forty-six percent of these children have home visitors who speak Spanish during visits.

Most Children Have Staff Who Use Their Home Language to Provide Services

Looking specifically at the language match between staff and families, we see that among all families, 86 percent (SE=3.04) of children have their home language used during home visits, and 93 percent (SE=1.75) have their home language used in the classroom (not shown). Of the Spanish-speaking families in home-based services, 78 percent (SE=6.89) have a Spanish-speaking home visitor. Of Spanish-speaking families in center-based services, 82 percent (SE=5.77) of children have a teacher or another adult in the classroom who speaks Spanish (not shown). These percentages are slightly lower than those found in 2009 and 2010. Note that examining the match between teachers' and families' language is complex because classroom teachers may have children speaking various languages within one classroom. In these cases, teachers may need another adult to help translate for or communicate with children.

Teachers and home visitors reported using a variety of strategies to communicate with families who speak a language the teacher or home visitor does not speak.[37] In these instances, less than two-thirds of children have teachers and home visitors who reported communicating with families only in English.[38] More teachers than home visitors reported using an informal interpreter to communicate with families (81 and 64 percent, respectively). Meanwhile, many children have a teacher (81 percent) or home visitor (78 percent) who reported using physical cues or hand gestures for communication with families.

Most Children Have a Teacher or Home Visitor with a College Degree and with Experience Working with Infants and Toddlers

The reauthorization of the Head Start Act requires that, nationally, all teachers serving children in center-based Early Head Start settings have at least a CDA (Child Development Associate) or equivalent state awarded credential by September 30, 2013 (U.S. Congress, H.R. 1429 Conference Report 2007; U.S. Department of Health and Human Services 2010). Home visitors in Early Head

[36] Adults in the classroom included in these estimates are the lead teacher, assistant teacher, classroom aide, or volunteer/non-staff.

[37] These findings apply to all the families served by the teachers and home visitors we interviewed who were working with any 3-year-old children in spring 2011 or 2012, not only to the children and families in the study sample.

[38] As discussed in Chapter III, 25 percent of children are spoken to at home in a language other than or in addition to English.

Start are required to have relevant experience and education, but there are no specific degree requirements. We stress that staff education reported at the child level should not be interpreted as reflecting the education of teachers or home visitors on average across all programs, or even on average within a program. The sampling design does not allow for reports at these levels.[39]

Many of the home-based and center-based children are being served by a staff member with at least a bachelor's degree (Table V.6). Specifically, more than half (59 percent) of children receiving home-based services have a home visitor with a bachelor's degree or higher and 26 percent have a home visitor with an associate's degree. Meanwhile, only one-third of children (37 percent) receiving Early Head Start center-based services have a teacher with a bachelor's degree or higher and 38 percent have a teacher with an associate's degree (AA). Among children whose teachers have at least an AA, 89 percent report that their field of study included early childhood education or child development. Similarly, among those whose home visitors have at least an AA, 89 percent studied early childhood education. Of teachers who have not earned a college degree, 88 percent have a CDA or a state-awarded preschool certificate; similarly, of home visitors who have not earned a college degree, 70 percent have a CDA or a state-awarded preschool certificate.[40]

On average, children in the home-based option have home visitors with more years of experience working with young children than do center-based children (see Table V.6). Children in the home-based option typically have home visitors with 11 years of experience working with infants and toddlers; center-based children's teachers average 8 years of experience (the medians are 10 and 7 years, respectively). Children have home visitors who have been working in Early Head Start for approximately 7 years and teachers who have been working in Early Head Start for 6 years (the medians are 7 and 5 years, respectively).

Teachers and Home Visitors Participate in a Number of Professional Development Activities

Children have teachers and home visitors who reported receiving substantial hours of staff training per year (see Table V.6). Children in the center-based option have teachers who reported attending an average of 64 hours of staff training annually, and those in the home-based option have home visitors who reported an average of 63 hours. At least three-quarters of children have a teacher (76 percent) or home visitor (79 percent) who receives both one-on-one and group supervision. Most commonly, supervision meetings are held at least once a month. Thirty-six percent of center-based children have a teacher with an assigned mentor/coach, as do 32 percent of home-based children and families. The frequency of meetings with these coaches varies, but at least three-quarters of children have a teacher (88 percent) or home visitor (84 percent) who meets with a coach at least once a month.

[39] Data reported here were collected in 2011 and 2012.

[40] Among children who have a teacher with less than a college degree, 82 percent (SE=4.75) had a CDA and 24 percent (SE=9.70) had a state-awarded preschool certificate. Among children who have a home visitor with less than a college degree, 58 percent (SE=16.3) had a CDA and 32 percent (SE=11.2) had a state-awarded preschool certificate.

Table V.6. Children's Teachers and Home Visitors Are Well-Educated and Experienced

Characteristics	Weighted Mean or Percentage (Standard Error)	
	Teachers	Home Visitors
Highest Level of Education (percentage)		
Less than high school	0.3 (0.27)	0.9 (0.89)
High school or equivalent	7.2 (4.12)	2.0 (1.27)
Some college but no degree	17.2 (2.92)	11.9 (3.13)
Associate's degree	38.3 (4.58)	26.4 (5.33)
Bachelor's degree	32.3 (5.28)	48.4 (6.08)
Graduate degree or higher	4.7 (1.79)	10.4 (3.48)
Field of Study Includes Early Childhood Education or Child Development (associate degree or higher; percentage)	88.9 (2.81)	88.5 (4.06)
Has a CDA or State-Awarded Preschool Certificate or License (less than a college degree; percentage)	87.8 (4.12)	69.6 (12.80)
Currently Enrolled in Child-care-related Training (percentage)	45.2 (4.52)	31.4 (4.38)
Years Teaching/Caring for Infants/Toddlers	7.9 (0.39)	10.6 (0.63)
Years Working in Early Head Start	5.6 (0.34)	6.7 (0.40)
Mean Hours of Staff Training per Year	63.8 (5.25)	62.5 (5.66)
Has Career or Professional Development Plan (percentage)	89.3 (2.88)	86.8 (3.93)
Supervision Meetings (percentage)		
One-on-one supervision	8.4 (2.24)	12.0 (3.20)
Group supervision	15.3 (4.29)	6.7 (2.52)
Both	75.9 (3.99)	78.7 (3.97)
None	0.4 (0.41)	2.6 (1.71)
Frequency of Supervision Meetings (percentage)		
At least once a month	81.4 (4.10)	86.0 (3.16)
Once every one to three months	11.4 (3.37)	11.4 (2.94)
One every four to six months	5.9 (2.20)	2.6 (1.37)
Once a year	1.3 (0.74)	0.0 (0.00)
Never	0.0 (0.00)	0.0 (0.00)
Has Mentor or Coach (percentage)	35.5 (4.62)	32.2 (5.48)
Frequency of Meetings with Mentor or Coach (percentage)		
Daily	2.5 (1.66)	6.2 (3.46)
Weekly	23.6 (6.91)	23.0 (7.03)
A few times a month	27.0 (7.63)	16.1 (4.92)
Once a month	34.8 (7.22)	38.3 (9.84)
More than once a year	7.3 (3.19)	14.4 (5.98)
Once a year	4.8 (2.63)	0.0 (0.00)
Never	0.0 (0.00)	2.0 (2.04)
Sample Size	**107-313**	**67-213**

Sources: Spring 2011 and 2012 Teacher and Home Visitor Interviews.

Note: Includes data on Early Head Start staff serving children in both cohorts. Because we did not sample teachers or home visitors, we must provide overall estimates as a percentage of children rather than as a percentage of teachers and home visitors.

CDA = Child Development Associate credential.

Children's Teachers and Home Visitors Report Positive Feelings About Their Current Job and Few Depressive Symptoms

Children's teachers and home visitors are generally positive about their profession and work in programs that offer a variety of benefits (see Table V.7). Most children have a teacher (86 percent) or home visitor (89 percent) who reported that he or she is very likely to stay in his or her job. At least 80 percent of children's teachers and home visitors reported receiving paid sick leave, paid holidays, paid vacations, retirement/pension plans, life insurance, health insurance, bereavement/family leave, and mileage reimbursement.

Table V.7. Teachers and Home Visitors Report Positive Feelings About Their Jobs, Many Benefits, and Few Mental Health Problems

	Weighted Percentage (Standard Error)	
Characteristics	Teachers	Home Visitors
Very Likely to Stay in Job (percentage)	86.1 (2.88)	88.7 (2.97)
Paid Sick Leave	90.7 (2.44)	96.5 (1.43)
Paid Holidays	95.3 (1.49)	94.1 (2.88)
Retirement/Pension Plan	80.7 (4.33)	91.1 (3.28)
Paid Vacations	86.2 (2.77)	92.7 (2.80)
Life Insurance	82.3 (3.45)	84.7 (3.69)
Paid Health Insurance	84.2 (2.72)	91.9 (2.64)
Dental Insurance	74.5 (4.18)	75.5 (5.67)
Paid Maternity Leave	70.5 (4.41)	70.1 (5.30)
Educational Stipends to Cover Workshops	65.6 (4.68)	69.8 (3.93)
Personal/Bonus Days	74.7 (4.10)	70.3 (5.96)
Bereavement/Family Leave	93.2 (2.18)	93.0 (2.19)
Mileage	84.5 (3.57)	92.1 (2.75)
Vision Care	69.2 (4.46)	69.7 (6.13)
Other	5.2 (1.72)	13.8 (2.86)
CES-D Short Form Scale Score	3.5 (0.33)	3.0 (0.36)
CES-D Short Form Categories (percentage)		
No/low number of symptoms	71.5 (3.86)	79.8 (3.60)
Mild symptoms	19.2 (3.67)	13.4 (3.07)
Moderate symptoms	6.6 (2.14)	4.5 (1.50)
Severe symptoms	2.6 (1.30)	2.3 (1.50)
Sample Size	**252-313**	**174-213**

Sources: Spring 2011 and 2012 Teacher and Home Visitor Interviews.

Note: The Center for Epidemiologic Studies Depression Scale (CES-D; Radloff 1977) uses 12 items to measure levels of depression among primary caregivers. Scores range from 0 to 36. Zero-4 = not depressed; 5-9 = mildly depressed; 10-14 = moderately depressed; 15 or more = severely depressed.

Includes data on Early Head Start staff serving children in both cohorts. Because we did not sample teachers or home visitors, we must provide overall estimates as a percentage of children rather than as a percentage of teachers and home visitors. Chapter II provides sampling information.

Existing research reports linkages between teacher psychological well-being and the quality of care children receive (Gerber et al. 2007, Vogel et al. 2011). Using the short form of the Center for Epidemiological Studies (CES-D; Radloff 1977; Ross et al. 1983), teachers and home visitors provide self-reports of their depressive symptoms that may influence the environment in Early Head Start classrooms, home visits, and staff interactions with children and families. Most children have teachers (91 percent) or home visitors (93 percent) who reported no or mild symptoms of depression (see Table V.7). About 7 to 9 percent of children have a home visitor or teacher who reported elevated (moderate or severe) numbers of symptoms.

Observed Quality in Early Head Start Programs

This section describes key aspects of children's home visits and classroom environments. We begin by offering a brief overview of the associations among observed quality, teacher and home visitor characteristics, and children's development.

For families served in Early Head Start home-based settings, home visiting is intended to support children's development, parenting outcomes, and the parent-child relationship. In the home-based setting there is a focus on process quality and not on structural quality, because structural factors such as teacher/child ratios and groups sizes do not apply. Process factors such as approaches and strategies used by home visitors vary, as does the efficacy of approaches in achieving the desired outcomes (Astuto and Allen 2009; Del Grosso et al. 2011; Paulsell et al. 2010; Roggman et al. 2008b; Sweet and Appelbaum 2004). Stronger effectiveness is likely when the quality of the home visit is high, including when the relationship between the home visitor and the family is strong (Paulsell et al. 2010). There is evidence that when visitors are matched to the family on characteristics such as ethnicity and language, parents are more engaged and retention is higher (Astuto and Allen 2009). Parents' engagement and involvement in a visit is also associated with the home visitor/parent relationship, and in turn, with children's vocabulary at 36 months (Roggman et al. 2008a).

In addition, qualitative research highlights the importance of home visitor conscientiousness (for example, honoring commitments to parents) and the match between home visitor and parent life experiences (Brookes et al. 2006). Others have found associations with child outcomes when home visits are child- rather than adult-focused (Raikes et al. 2006). Findings suggest that the content of home visits could be important for enhancing desired child and family outcomes.

Turning to the classroom setting, existing research emphasizes links between both process factors (including teacher behavior, teacher-child interactions, and quality of instruction) as well as structural factors (including child-teacher ratios, group sizes, and teacher education) and child developmental outcomes (Love et al. 2005; Phillips et al. 2000; Phillipsen et al. 1997). For example, modest associations between observed classroom quality and child outcomes for low-income infants, toddlers, and preschoolers have been found in individual studies and meta-analyses (Burchinal et al. 1996, 2008a, 2009). Sensitive and responsive interactions with teachers are particularly important for children's learning and social-emotional development (NICHD Early Child Care Research Network 1996, 1997, 1998, 2006; Whittaker and Harden 2010). In fact, process characteristics, such as sensitive and stimulating interactions with teachers, are associated with prekindergarten children's language, preacademic, and social skills (Burchinal et al. 2008b). Considering structural features, the positive association between a low adult-to-child ratio and child outcomes has been well documented (Burchinal et al. 1996; Scarr et al. 1994; Whitebook et al. 1989).

Box V.1 provides information on the measures used to assess the key aspects of quality for both home- and center-based services for children in the Baby FACES study.

Box V.1. Measures of Home Visit and Classroom Quality

To assess key aspects of the quality of both home visits and center-based classrooms, field staff conducted structured observational assessments of home visits and classrooms of 3-year-old children. We observed home visits for those receiving child development services primarily through home visits and observed center-based classrooms for infants and toddlers receiving child development services primarily in a center-based setting. We observed one home visit per home visitor (not one per child).[41] For home visits, we used the Home Visit Rating Scale-Adapted (HOVRS-A) (Roggman et al. 2009) and its manual (Hallgren et al. 2009), an adaptation of the HOVRS (Roggman et al. 2006b). For classroom observations, we used the Classroom Assessment Scoring System-Toddler (CLASS-T) (Pianta et al. 2010) for classrooms serving 3-year-old children.

HOVRS-A, originally developed for training Early Head Start home visitors, is based on a theoretical perspective of an optimal model of home visiting. In this model, home visitors facilitate developmentally appropriate parenting behaviors and build upon parents' skills and resources to support child development (Roggman et al. 2008b). This approach focuses more on the parent-child interaction and less on one-on-one interaction with either the parent or child. Higher scores on the HOVRS-A have been associated with higher scores on a measure of the quality of the home environment (Roggman et al. 2006a), and the quality of the home environment has been found to mediate children's language development (Tamis Le-Monda et al. 2005).

HOVRS-A consists of seven items, which can be combined to form a total score and two subscale scores: Visitor Strategies (4 items) and Visitor Effectiveness (3 items). Visitor Strategies items include (1) the home visitor's responsiveness to the family, (2) the home visitor–family relationship, (3) the home visitor's facilitation of parent-child interaction, and (4) the home visitor's nonintrusiveness. Visitor Effectiveness includes (1) parent-child interaction during the visit, (2) parent engagement, and (3) child engagement. Items on HOVRS-A are rated from 1 to 5, based on indicators defined for each item and with anchor ratings of 1 (minimal), 3 (adequate), and 5 (good practice). A 3 rating indicates that the observer saw a sufficient level of indicators of the model of home visiting that aims to facilitate parent-child interaction. Because of the clear theoretical underpinnings of this instrument, home visiting programs that adhere to alternative models of home visiting and stress different types of behaviors (such as home visitor–child interactions) will not score as highly on the HOVRS-A.

During observations of home visits, field staff also collected data on the content and characteristics of the home visit, including topics (such as the child's health and development, parenting, the parent's health and well-being, parent employment and education, and community services); activities (including assessment, provision of information, goal-setting, and crisis intervention); and structure (for example, participating children and adults and languages used) (Boller et al. 2009).

Classroom observations of 3-year-olds were conducted using the CLASS-T (Pianta et al. 2010), a downward extension of the Pre-K CLASS (Pianta et al. 2008), which focuses on teacher-child interaction quality in toddler child care classrooms. The CLASS-T measures process quality along eight dimensions (Positive Climate, Negative Climate, Teacher Sensitivity, Regard for Child Perspectives, Behavior Guidance, Facilitation of Learning and Development, Quality of Feedback, and Language Modeling). The dimensions exist within two broader domains: Emotional and Behavioral Support and Engaged Support for Learning. Dimensions are defined by observable indicators along a seven-point scale, with ratings reflecting scores in the low (1-2), mid (3-5), and high (6-7) ranges. Appendix D presents additional information, including procedures for scoring the CLASS-T and findings of principal components factor analysis.

Classroom observations for 3-year-olds also included counts of infants and toddlers and the adults caring for them, which we used to compute child-adult ratios and group sizes.

For the HOVRS-A, ITERS-R, and CLASS-T, observers look for evidence of specific indicators as they rate each item. The unweighted means, standard deviations, and ranges for scores at age 3 (that is, HOVRS-A and CLASS-T scales) are presented below.

Spring 2011 and 2012 HOVRS-A and CLASS-T Scores, Unweighted

Domain	Mean	Standard Deviation
HOVRS-A Overall Quality	3.4	0.76
Visitor Strategies Quality	3.2	0.90
Visitor Effectiveness Quality	3.8	0.83
CLASS-T Emotional and Behavioral Support	5.3	0.89
Positive Climate	5.6	1.15
Negative Climate	1.3	0.71
Teacher Sensitivity	4.8	1.04
Regard for Child Perspectives	4.7	1.04
Behavior Guidance	4.7	1.16
CLASS-T Engaged Support for Learning	3.2	1.16
Facilitation of Learning and Development	3.7	1.13
Quality of Feedback	3.1	1.19
Language Modeling	2.9	1.32
Sample Size	**181-314**	

Source: Spring 2011 and 2012 Home Visit Observation and Classroom Observation.

Note: Scores are reported only for children at age 3. The CLASS-T Emotional and Behavioral Support composite score includes reverse-coded values on Negative Climate.

HOVRS-A=Home Visit Rating Scale-Adapted; CLASS-T=Classroom Assessment Scoring System-Toddler.

Children and Families in the Home-Based Option Participate in a Variety of Activities During Visits

Early Head Start performance standards require that families receiving home-based services receive weekly home visits that last 90 minutes (see http://www.ehsnrc.org/PDFfiles/EHS-Home-AdminChk.pdf). Based on our observations, the length of the home visits provided to sample children and their families is just under 90 minutes, on average (mean = 80 minutes; range = 35 to 150 minutes; see Table V.8). We emphasize this does not necessarily imply that 80 minutes is the typical length of all home visits; it is only the average of those we were able to observe. It is also possible that the presence of an observer during a home visit could alter the length, content, or dynamic of the interactions in unknown ways.

During home visits, home visitors are encouraged to meet family needs by engaging those present during the visits. For example, if the family has more than one child, the home visitor may be able to address the needs of both the child in the program and the other children in the family by engaging parents in activities with all of the children or by answering parents' questions about siblings and their development. On average, during visits we observed one additional child (other than the focal child) and one adult present and engaged with the home visitor. Eighty-five percent (SE=3.84) of the time, the child's mother or female guardian was the adult present during the home visit.

[41] We based our decision to observe one visit per home visitor rather than per child on two considerations. The first was the logistical difficulty of scheduling an observation of each child's home visit in the site visit week (five days). The second was that prior research suggested that home visitors tend to have low intra-visitor variability; in other words, home visitors tend to provide home visits of a consistently similar quality (personal communication with Lori Roggman, 2008).

Table V.8. Children and Families in the Home-Based Option Participate in a Variety of Activities During Observed Visits

Characteristics	Weighted Mean or Percentage (Standard Error)
Length of Home Visit (minutes)	80.3 (2.99)
Number of Children Other than Focus Child Participating in Visit	0.5 (0.09)
Number of Adults Other than Home Visitor Participating in Visit	0.8 (0.09)
Home Visit Conducted in (percentage)	
English	81.9 (3.71)
Spanish	44.0 (7.91)
Other Language	4.7 (2.94)
If Home Visit Conducted in Language Other than English, Interpreter Used (percentage)	25.2 (7.02)
Proportion of Home Visit Time per Type of Activity (percentage)	
Child-focused activities	57.0 (2.92)
Parent-family-focused activities	12.4 (1.15)
Parent-child-focused activities	13.6 (1.60)
Staff-family relationship-building activities	15.6 (2.99)
Crisis management activities	1.2 (0.34)
Activities During Home Visit[a] (percentage)	
Child/parent observation/assessment	32.1 (4.66)
Evaluation/feedback on parent-child interactions	22.5 (4.16)
Provision of education and/or information	67.4 (5.02)
Problem solving	22.9 (5.48)
Goal setting/planning	42.0 (5.31)
Crisis intervention	3.8 (2.13)
Model or demonstrate interaction with child/facilitate parent-child interaction	43.8 (5.65)
Observation of caregiver-child interactions	22.4 (4.33)
Provision of emotional support to parent	15.0 (3.00)
Play	90.7 (2.56)
Other	8.1 (4.22)
Alignment of Home Visit Activities with Planned Activities[b]	4.4 (0.11)
Sample Size	**171-178**

Source: Spring 2011 and 2012 Home Visit Observations.

Note: Because we did not sample teachers or home visitors, we must provide overall estimates as a percentage of children rather than as a percentage of teachers and home visitors.

[a]Activity categories do not sum to 100 because more than one activity could occur during the home visit.
[b]Rated on a scale of 1 to 5, with 5 indicating "very well aligned" with planned activities.

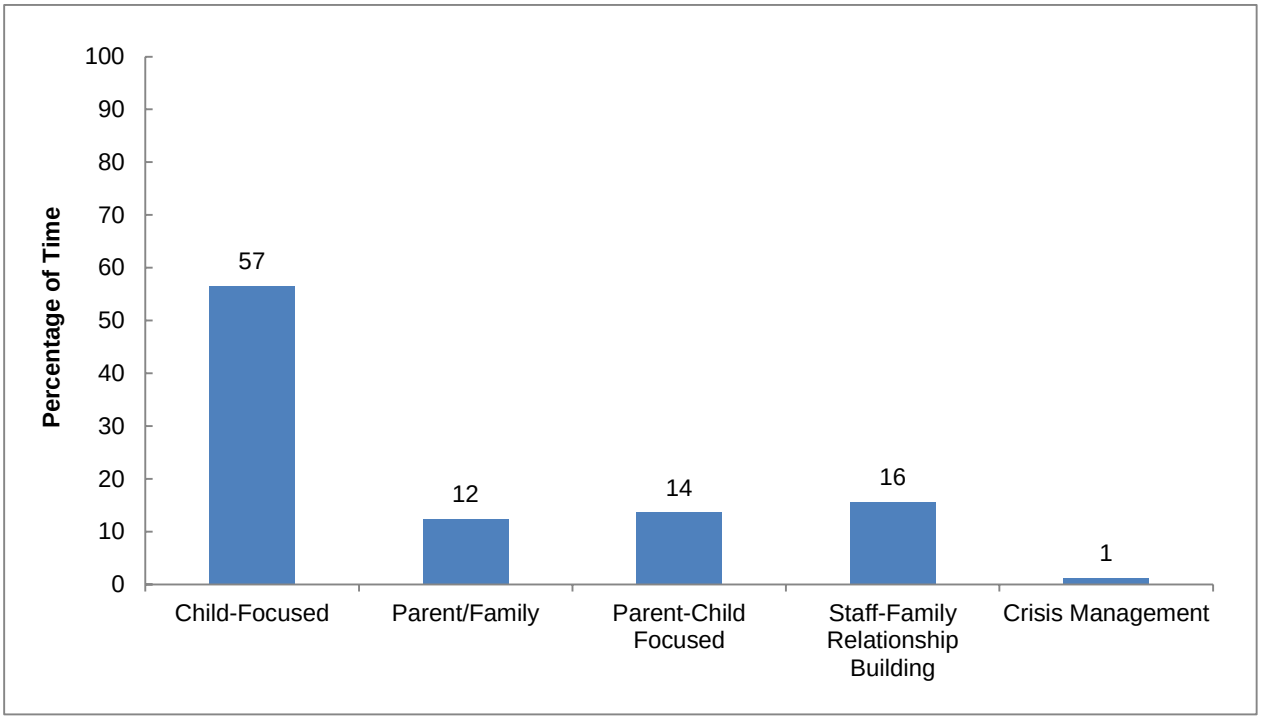
Programs strive to provide home visiting services in the language parents and children are most comfortable speaking. Among the visits we observed (one per home visitor), most (82 percent) were conducted in at least English; however, 44 percent were conducted in at least Spanish, and 5 percent were conducted in at least a language other than English or Spanish.[42] Five percent of observed visits were conducted in more than one language. For approximately two-thirds of children from homes in which Spanish is the primary language (63 percent), the home visit is conducted in Spanish.[43] Interpreters are used in 25 percent of all visits in which a language other than English is used.

[42] Note that the source for this information differs from earlier reports of the percentage of children having a home visitor speaking a language other than English. That information is drawn from the home visitor interview, while the information noted here is drawn from the home visit observation.

[43] The information presented here is drawn from multiple questions about family language and service delivery approaches. Thus, the responses do not necessarily total 100 percent.

Based on our observations, the largest proportion of time (57 percent) during a home visit is spent on child-focused activities (Figure V.1). Additionally, 16 percent of the time is focused on staff-family relationship building activities, 14 percent on parent-child activities, and 12 percent on parent and family activities. About 1 percent of the home visit time is focused on crisis management.

Figure V.1. Time Allocated to Types of Activities, Age 3



Source: Spring 2011 and 2012 Home Visit Observation.

Notes: N=178; 1-year-old and Newborn Cohorts at age 3.

Home visits most commonly involve play (91 percent) and provision of education or information (67 percent). Less than half of children have home visits that address goal setting and planning (42 percent). Visits with families also include time spent modeling or facilitating parent-child interactions (44 percent), child/parent observation and assessment (32 percent), and observation of parent-child interactions (22 percent). About one-quarter of children have visits that include feedback on parent-child interactions (23 percent) and problem-solving activities (23 percent). After the visits were completed, home visitors reported that observed visits were highly aligned with the activities they had planned and that they were able to cover the topics and perform the activities on which they had set out to work.

Home Visitor Caseloads and Frequency of Visits Are Within the Performance Standards and Professional Recommendations

Program directors report that home visitors have an average caseload of 10 families (Table V.9). This number falls within performance standards (10 to 12 families per home visitor; see EHSNRC 2009). For home-based services, programs are required to offer weekly home visits. Program directors report that most families receive home visits fairly frequently. Program directors report that 71 percent of all enrolled families receive visits four or more times a month, and another 18 percent receive them three times a month. Eleven percent receive up to two visits per month. Chapter VII discusses service offerings to study families based on weekly reports completed by teachers and home visitors.

Table V.9. Director-Reported Home Visit Caseload and Frequency Meet Standards

Characteristics	Weighted Mean or Percentage (Standard Error)
Mean Number of Families per Home Visitor	10.2 (0.27)
Percentage of Home-Based Families Receiving Home Visits	
Four or more times a month	71.4 (3.74)
Three times a month	17.6 (2.71)
Two times a month or less	11.0 (2.19)
Sample Size	**77**

Source: Spring 2011 Program Director Interview.

Note: Items asked only of program directors reporting a home-based option.

Most Children and Families in the Home-Based Option Receive Home Visits of Mid-range Quality

On the HOVRS-A Visitor Strategies subscale, home visits scored in the 3-point range, on average (mean=3.2; see Table V.10). This subscale includes four items that capture the home visitor's interactions and relationship with the parent and child. About two-thirds of families have visits scoring between 2 and a 4 on this subscale (67 percent). Scores in the 3-point range on this scale indicate that home visitors occasionally use strategies in their interactions that demonstrate responsiveness and that help develop relationships with parents and children. It also means that the home visitor occasionally tries to facilitate and reinforce parent-child interactions rather than interacting solely with the parent or child, and occasionally guides (rather than controls) aspects of parent-child interactions. Ten percent of families have a visit scoring lower than 2 on this subscale, and 23 percent have a score of at least 4. Less than 1 percent have a visit with a score of 5 (see Figure V.2).

On the HOVRS-A Visitor Effectiveness subscale, which captures the home visitor's effectiveness in involving and engaging the family, the average subscale score is 3.8. About half of families have a home visit scoring a 4 or higher on this subscale (54 percent), and 8 percent score a 5 (see Figure V.2). These average scores suggest that families have home visitors who are relatively effective at engaging parents and children with each other and with the activities of the home visit. On this subscale, visits score highest in child engagement (mean of 4.7) compared with the other two items included in this subscale. Therefore, during home visits, children at least occasionally interact with the parent or home visitor and demonstrate interest in home visit activities. Scores on both subscales indicate that the home visitor might benefit from coaching. Scores and patterns on the HOVRS-A are similar to findings at age 2.

Table V.10. Most Children and Families in the Home-Based Option Receive Visits of Mid-range Quality

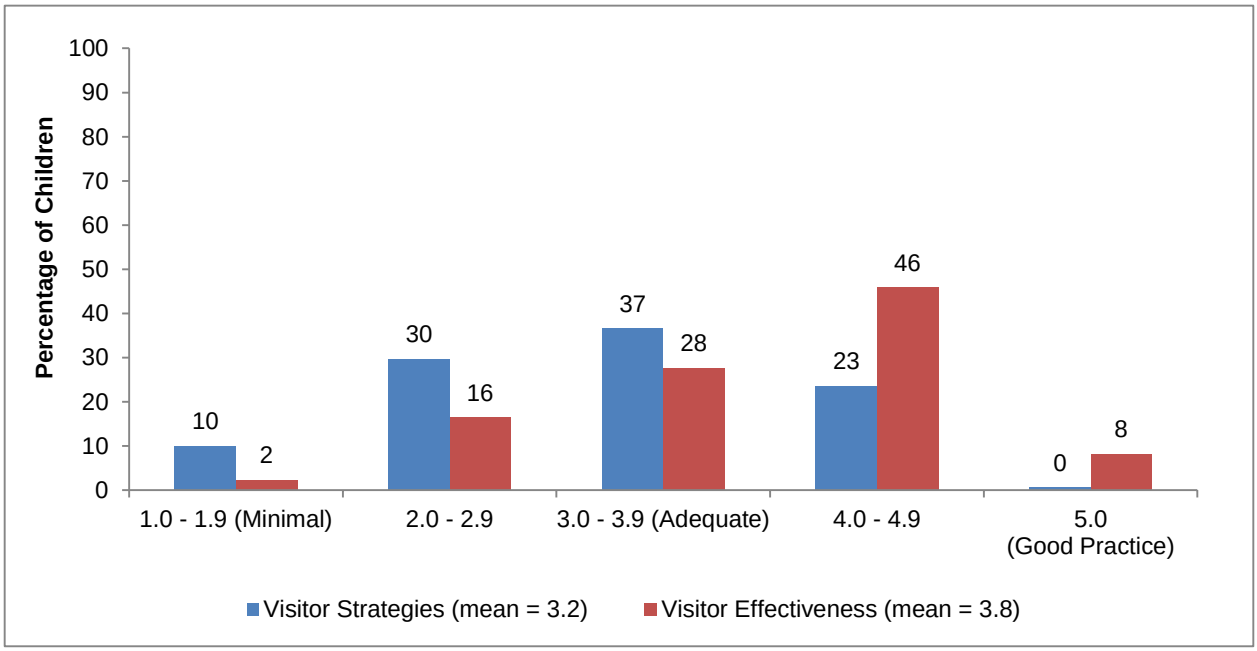
Scales	Weighted Mean (Standard Error)
HOVRS-A Visitor Strategies Quality	3.2 (0.08)
Responsiveness to family	3.1 (0.08)
Relationship with family	4.1 (0.07)
Facilitation of parent-child interaction	2.8 (0.10)
Nonintrusiveness	2.7 (0.15)
HOVRS-A Effectiveness Quality	3.8 (0.11)
Parent-child interaction	3.4 (0.16)
Parent engagement	3.4 (0.16)
Child engagement	4.7 (0.08)
Observer Rating of Visit Quality[a]	3.1 (0.12)
Sample Size	**169-178**

Source: Spring 2011 and 2012 Home Visit Observations.

HOVRS-A = Home Visit Rating Scale-Adapted.

[a]As part of the home visit observation, field staff provided an overall rating of the quality of the home visit based on the content of the visit and the quality of the relationship and interactions between the parent and home visitor. Ratings range from poor (1) to excellent (5).

Figure V.2. Observed Home Visits Typically Fall in the Mid-range of Quality



Source: Spring 2011 and 2012 Home Visit Observation.

Note: N = 178; 1-year-old and Newborn Cohorts at age 3. We point out that the nature of the scoring rubric is such that each of the anchor points includes a range of scores (for example 2.0 to 2.9, etc.), with the exception of 5, the highest point. We elected not to collapse scores to create a range because that would entail a loss of precision and useful information about the highest end of the scale.

HOVRS-A = Home Visit Rating Scale-Adapted.

The next sections describe classroom quality, beginning with aspects of structural quality, then turning to process quality measures. All information is based on observations of classrooms that included 3-year-old study children.

Classrooms Offer Group Sizes and Ratios Within the Performance Standards and Professional Recommendations

On average, children are in Early Head Start classrooms with observed group sizes of 6.1 children and child-teacher ratios of 2.7 children per teacher (Table V.11). These numbers fall within performance standards (4 children per adult and a maximum group size of 8; see ACF 2009a).

Table V.11. Most Children Are in Classrooms of Moderate Quality

	Weighted Mean (Standard Error)
Group Size	6.1 (0.14)
Child/Adult Ratio	2.7 (0.09)
CLASS-T Emotional and Behavioral Support	5.3 (0.09)
Positive Climate	5.5 (0.13)
Negative Climate	1.4 (0.07)
Teacher Sensitivity	4.8 (0.10)
Regard for Child Perspectives	4.7 (0.10)
Behavior Guidance	4.7 (0.11)
CLASS-T Engaged Support for Learning	3.3 (0.13)
Facilitation of Learning and Development	3.7 (0.13)
Quality of Feedback	3.1 (0.12)
Language Modeling	2.9 (0.15)
Sample Size	**297-304**

Source: Spring 2011 and 2012 Classroom Observations.

CLASS-T = Classroom Assessment Scoring System-Toddler.

Nearly all Baby FACES children (98 percent) are in classrooms with observed group sizes of 8 or fewer children, and nearly all are in classes with ratios of 4 to 1 or better (98 percent). This ratio is slightly smaller than we observed at 24 months in the EHSREP (3.5 children per adult).

Relatively Wide Age Ranges Are Common in Classrooms

Many classrooms serving 3-year-olds that we observed included a fairly wide range of ages, with an average span of almost 13 months between the youngest and oldest child (ranging from 0 months to 33 months). To understand the proportion of "mixed-age" classrooms in the sample, we examined the percentage of classrooms of 3-year-olds that included children who were 24 months or younger and/or 48 months or older. According to this criterion, 46 percent of the classrooms serving 3-year-olds that we observed are mixed-age. Within these mixed-age classrooms, the average age span between the youngest and oldest child is 19 months (ranging from 3 months to 33 months), as compared with a 7-month average age span (ranging from 0 months to 19 months) among children in non-mixed classrooms.

Most Children in Center-Based Programs Are in Classrooms of Moderate Quality

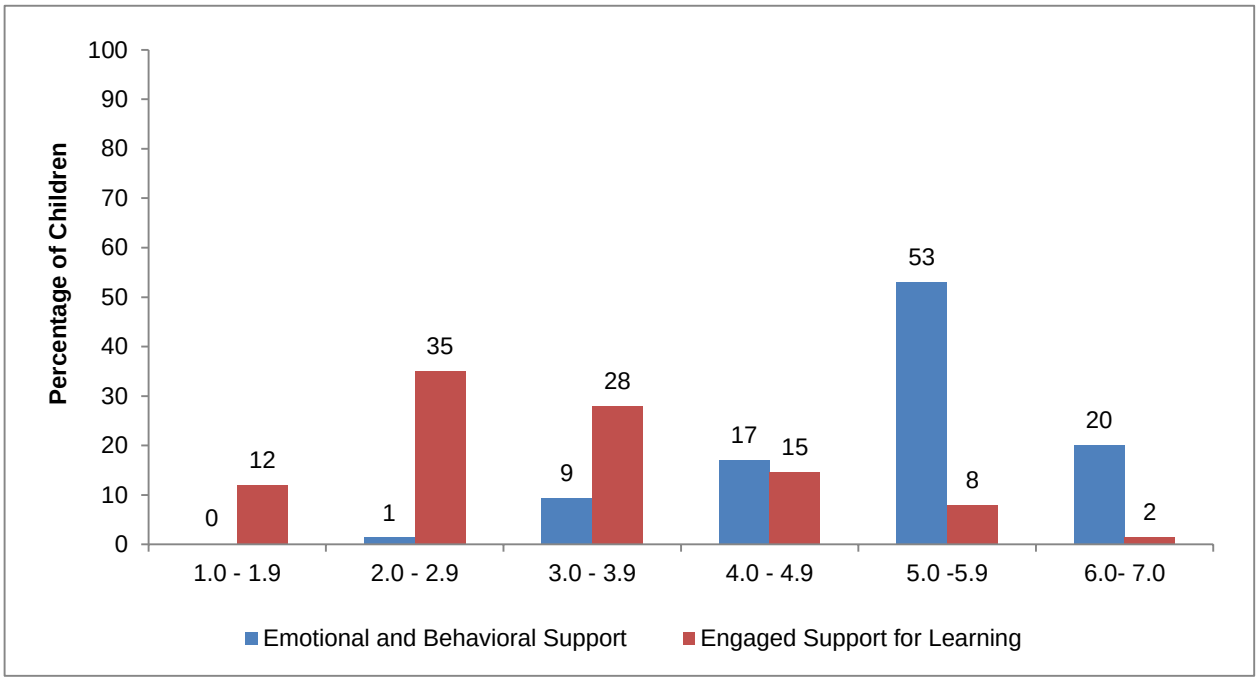
For observations of process quality in center-based classrooms, we used the Classroom Assessment Scoring System-Toddler (CLASS-T) for classrooms serving 3-year-old children. The CLASS-T is a new measure that rates teacher-child interactions and is intended for use with children ages 15 to 36 months. In the Baby FACES sample, we find that 3-year-old children are in classrooms that score a mean of 5.3 and 3.3 (out of 7) in the domains of Emotional and Behavioral Support and

Engaged Support for Learning, respectively (Table V.11).[44] Although scores in the domain of Engaged Support for Learning are lower, on average, than those observed in the domain of Emotional and Behavioral Support, scores across both domains represent ratings in the moderate range of quality (scores of 3 to 5). Classrooms are strongest in the area of Emotional and Behavioral Support, including the dimensions of Positive Climate (5.5), Teacher Sensitivity (4.8), Regard for Child Perspectives (4.7), and Behavior Guidance (4.7). Classrooms were rated in the low range on Negative Climate (1.4), indicating that interactions characterized by negativity were infrequently observed. Classrooms scored lowest in the area of Engaged Support for Learning, including the dimensions of Facilitation of Learning and Development (3.7), Quality of Feedback (3.1), and Language Modeling (2.9). In the pilot study of the adapted toddler measure (Thomason and LaParo 2009), Language Modeling was likewise rated the lowest of all the observed dimensions (1.9). Findings on Engaged Support for Learning are similar to those in studies with older preschool children using the preschool CLASS (for example, FACES, Aikens et al. 2010; C-PEP, Ross et al. 2008) and suggest that providers face challenges in their attempts to offer high-quality instructional support to children, including facilitating activities that support children's learning and development, providing individualized feedback to promote children's understanding of concepts, providing opportunities for children to use language, and engaging in conversations that extend children's language skills. In the domain of Emotional and Behavioral Support, findings are similar to those found in observations of classrooms serving 2-year-old children in Baby FACES. However, average scores in the domain of Engaged Support for Learning are up to half a point lower than those observed at age 2; in the area of Language Modeling in particular, mean scores are in the low range. Across all the CLASS-T dimensions, these Early Head Start classroom scores compare favorably with those reported by the CLASS-T authors across a number of reports (La Paro et al. 2014; Thomason and La Paro 2009).[45]

[44] As noted in Box V.1, for classroom observations, we used the CLASS-T for children in the Newborn Cohort who were 3 years old in spring 2012 and for children in the 1-year-old Cohort who were 3 years old in spring 2011.

[45] The measure used by Thomason and La Paro (2009) was a downward extension of the CLASS (Pianta et al. 2008) that included only six of the eight component dimensions: Positive Climate, Negative Climate, Teacher Sensitivity, Regard for Child Perspectives, Behavior Guidance, and Language Modeling. Consequently, the study authors did not report findings for the dimensions of Facilitation of Learning and Development and Quality of Feedback, or the resulting composite domain score derived from these dimensions. The Emotional Climate domain score reported by the authors is similar to the composite measure of Emotional and Behavioral Support derived in Baby FACES; a key dissimilarity is the inclusion of Behavior Guidance in the Emotional and Behavioral Support composite score. (See Appendix D for findings of principal components factor analysis using the Baby FACES study sample.)

Figure V.3. Most Children Are in Classrooms in the Mid-Range of Quality



Source: Spring 2011 and 2012 Classroom Observation.

Note: The overall mean score was 5.3 for Emotional and Behavioral Support and 3.3 for Engaged Support for Learning.

Sample size: N = 312. 1-year-old and Newborn Cohorts at age 3.

CLASS-T = Classroom Assessment Scoring System-Toddler.

Using the developer-provided definitions of the CLASS-T dimension scores, almost all children in Baby FACES are in classrooms rated as falling in the mid-to-high range in the domain of Emotional and Behavioral Support (99 percent; Figure V.3). Twenty percent are in classrooms rated as 6 or higher. In contrast, only 53 percent of children are in classrooms receiving scores in the mid-to-high range for Engaged Support for Learning. About half (47 percent) of all children are in classrooms scoring in the low range, with only 2 percent of children in classrooms rated as 6 or higher.

Parents and Staff Report Positive Relationships with One Another

Communication between parents and teachers or home visitors, as well as agreement between parents and these staff on child-rearing philosophy, has been related to child outcomes. Particularly with home visiting services, the quality of the relationship between the home visitor and the parent may influence the effectiveness of care and the extent and quality of parent engagement and involvement (Korfmacher et al. 2007, 2008; Roggman et al. 2008b). Accordingly, we included items from the Parent-Caregiver Relationship Scale (PCRS; Elicker et al. 1997) in Baby FACES to assess the quality of this relationship (see Box V.2).

Box V.2. Measuring the Parent-Staff Relationship

The Parent-Caregiver Relationship Scale (PCRS; Elicker et al. 1997) measures the perceived relationship between the parent and the Early Head Start staff member who delivers the primary service to the child and family (that is, the teacher or the home visitor). Items capture important dimensions of the parent-staff relationship, including trust and confidence, communication, respect and acceptance, caring, competence and knowledge, partnership and collaboration, and shared values. The spring 2011 and 2012 Baby FACES instruments included items across these dimensions, adapted for use with home visitors from the original version developed primarily for center-based teachers. Parents and staff rated items on a scale from 1 to 5 (that is, from strongly disagree to strongly agree). Scale scores represent the average across a subset of these items (six and seven items for staff and parents, respectively). Appendix C presents additional information on the reliability of these scales in spring 2011 and 2012.

Spring 2011 and 2012 Staff-Parent Relationship Quality Scores, Unweighted

Domain	Mean	Standard Deviation	Range
Staff-Parent Relationship Quality Scores for Children Receiving Services by Home Visits			
Parent report[a]	4.7	0.53	1.0 – 5.0
Home visitor report[b]	4.4	0.65	1.0 – 5.0
Staff-Parent Relationship Quality Scores for Children Served in Centers			
Parent report[c]	4.6	0.62	1.0 – 5.0
Teacher report[d]	4.2	0.83	1.0 – 5.0
Sample Size	**169-305**		

Source: Spring 2011 and 2012 Staff-Child Report and Parent Interview.

Note: Scores are reported here only for staff and parents of children at age 3.

[a]Sample size of 169.
[b]Sample size of 206.
[c]Sample size of 238.
[d]Sample size of 305.

PCRS = Parent-Caregiver Relationship Scale.

On average, parents agree or strongly agree with positive statements about the quality of relationships with their home visitor or teacher (mean=4.6 and 4.7 for those receiving home visits and those with children in centers, respectively; see Table V.12). For example, they typically agree or strongly agree with statements such as, "If there is a problem, my child's teacher/home visitor and I always talk about it soon" and "I feel that my child's home visitor/teacher genuinely cares for [my child]." As in spring 2009 and 2010, teachers and home visitors express similar positive attitudes about their relationships with children's parents (mean=4.1 and 4.4, respectively).

Table V.12. Parents and Staff Report Positive Relationships with One Another

Characteristic	Weighted Mean (Standard Error)
Staff-Parent Relationship Quality Score for Children Served in Centers	
Parent Report	4.6 (0.04)
Teacher Report	4.1 (0.06)
Staff-Parent Relationship Quality Score for Children Served by Home Visits	
Parent Report	4.7 (0.06)
Teacher Report	4.4 (0.05)
Sample Size	**150-248**

Sources: Spring 2011 and 2012 Parent Interview and Staff-Child Report.

Relationship quality is not associated with observations of home visit or classroom quality, as measured by the HOVRS-A and CLASS-T, respectively. Correlations between parent ratings of the relationship and staff ratings were likewise not statistically significant. In a study of 217 parents and caregivers, Elicker and colleagues (1997) also found that correlations among the parent and caregiver scales were not significant, suggesting that parent-caregiver reports were incongruent. The authors note that typically on measures of perceived relationships, staff less frequently report positive ratings of parents than parents report of staff, with staff ratings varying with the demographic characteristics of parents (for example, age, income, and marital status).[46]

Measuring Program Implementation

Program implementation has increasingly been recognized as a critical component for understanding program effects (or the lack thereof) (Damschroder and Hagedorn 2011; Durlak and DuPre 2008). In the Early Head Start context, importance of fully implementing the performance standards was demonstrated in the EHSREP when programs that were fully implemented early achieved more and larger impacts on children's development and parenting (ACF 2002a). As a result, one of the goals of Baby FACES is to better understand how current programs are implemented. Box V.3 provides background on implementation measurement in the EHSREP and describes the approaches we have adopted in Baby FACES.

The implementation scores for 2009, 2010, and 2011 yielded ratings that can range from 1 to 4, where 1 represents low or minimal implementation, 2 represents moderate implementation, 3 represents full implementation, and 4 represents enhanced implementation. Each of the cornerstones (see Figure I.1) is comprised of multiple component items. In turn, a cornerstone rating is the average of its components. The overall implementation ratings in 2010 and 2011 are based on an approach that requires all cornerstones to be fully implemented in order to be rated as fully implemented overall.[47] The rating scales in Appendix E include the criteria for assigning ratings to each component of each cornerstone.[48]

[46] Correlations between staff relationship ratings and parent characteristics are small in magnitude (approximately .11 to .15) but statistically significantly and positively associated with age, income, and married marital status.

[47] We use sample weights accounting for the sample design and survey nonresponse, so findings pertain to the Early Head Start programs nationally.

[48] Implementation ratings in 2009 were different in format and mode of administration than in 2010 and 2011 and scores were also calculated somewhat differently. Overall implementation scores are the average of the cornerstone ratings, with child development counted twice to increase the contribution of this cornerstone to the overall ratings.

Box V.3. Measuring Implementation in the EHSREP and in Baby FACES

In the EHSREP, researchers developed detailed implementation rating scales based on the Head Start Program Performance Standards and constructed rating scales for each program cornerstone (child development, family development, staff development, community building) and for management systems (see Figure I.1). The data for these ratings was collected in multiday site visits and examination of program documents. Site visitors and two other team members completed the rating scales and reached consensus in the event of discrepant ratings (ACF 2002b).

The ratings were later validated through their associations with program impacts. Analysts grouped programs according to their pattern of implementation across two time periods, two years apart: early implementers (programs that were rated as fully implemented at both time points), later implementers (programs that were fully implemented at the second time point, but not the first), and incomplete implementers (programs that were not fully implemented at either point). Program impacts for children at age 3 (the end of Early Head Start eligibility) were strongest for early implementers (ACF 2002a).

The approach to measuring implementation in Baby FACES was guided by the process used in the EHSREP, although administered differently. Because intensive data collection during in-person site visits was not feasible for Baby FACES, the research team explored alternative approaches to gathering information about program implementation. In 2009, researchers asked program directors to rate their own program using as a guide a scale with comprehensive descriptors for each anchor point, which was heavily based on summary ratings used in the EHSREP.

In 2010 and 2011 we tried a different approach to measuring implementation. The team developed items with concrete response categories that tracked to each element in the rating form.[49] That approach had the advantage of being less subjective, first because program directors could not see the ratings they were awarding themselves and second, because it allowed us to ensure that all requirements for a given anchor were met. Program directors responded to these survey items during the program director interview, and these were then coded by the analysis team into overall ratings of implementation. To do this, we had to adapt the rating scales and include only information that could be obtained in a survey. We therefore developed schemes to compute ratings. To the extent possible, we simulated the process of creating implementation ratings in the EHSREP, and we consulted with OHS and our federal project officer to ensure that the rating schemes were reasonable and sensible. Appendix E contains the adapted rating scales we used in this follow-up.

Nearly Half of Programs Are Fully Implemented Overall

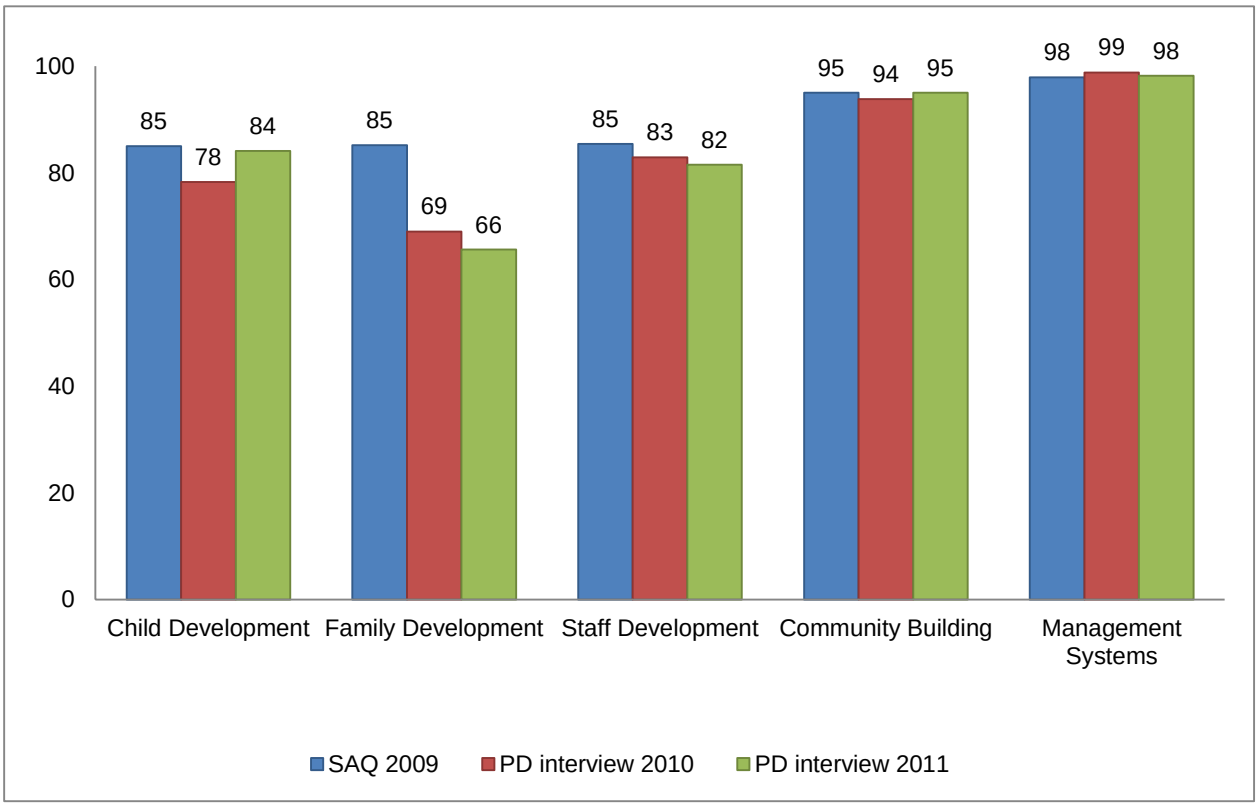
Just under half of the programs are rated as fully implemented (full implementation in all cornerstones and the management area) overall in 2010 and 2011 (41 percent in 2010 and 45 percent in 2011). The remaining programs did not implement all components of at least one cornerstone or management area. Of note, the degree of program implementation varies by the mode and manner of measurement. The current findings differ a great deal from survey data collected in 2009 where about 72 percent of programs were fully implemented according to self-ratings by directors.

[49] For example, in the 2009 self-rating form, the anchor for a rating of enhanced implementation for "parent involvement in child development services" reads, "At least one parent in many enrolled families participates in planning and delivering child development services. Of those families with a father or father figure, many fathers participate in child development services." The 2010/2011 interview broke down this description into two items about proportions of parents involved in child development service planning as well as father involvement in planning. Further, it gave ranges to quantify the proportions involved. The analysis team, in consultation with OHS and other experts, developed cut points to determine the level of involvement that would be needed for each rating.

Implementation Varies by Cornerstone

The overall implementation ratings mask variability within some cornerstones. In general, management systems and community building were rated as highly implemented across nearly all programs and years. Programs were, however, more variable in implementing child, and family development cornerstones in a given year and over time, and these are presented below (Figure V.4). Table V.13 shows variability within cornerstone elements.

Figure V.4. Percentage of Programs Fully Implemented by Cornerstone 2009 – 2011



Source: Spring 2009 Program Director Self-Administered Questionnaire, 2010, and 2011 Program Director Interviews.

Note: Program director self ratings were collected in 2009 and survey items administered in the program director interview were rated by staff in 2010 and 2011.

Sample size: N = 89 programs in 2010 and 2011. Sample size in 2009 was 86, and missing data are imputed in the figure.

Table V.13. Average Scores of Child and Family Development Cornerstones Across Years

	Weighted Means (Standard Error)		
	2009[a]	2010	2011
Child Development Cornerstone			
Frequency of child development services	4.1 (0.13)	3.3 (0.06)	3.3 (0.04)
Developmental assessments	4.8 (0.06)	3.5 (0.04)	3.6 (0.03)
Health services	4.9 (0.04)	2.9 (0.02)	3.0 (0.02)
Child care	3.5 (0.17)	2.8 (0.06)	2.6 (0.08)
Parent participation in child development services planning	4.0 (0.10)	n.a.	n.a
Individualization	4.9 (0.03)	3.8 (0.08)	3.8 (0.08)
Group socializations	3.5 (0.15)	2.6 (0.04)	2.6 (0.05)
Family Development Cornerstone			
Individualized Family Partnership Agreements (IFPAs)	4.4 (0.11)	3.2 (0.06)	3.2 (0.06)
Availability of services	4.5 (0.08)	3.3 (0.06)	3.3 (0.05)
Frequency of family development services	4.5 (0.09)	n.a	n.a
Parent involvement in program activities	3.8 (0.11)	3.0 (0.06)	2.8 (0.07)
Sample Size	**79-86**	**87-89**	**86-89**

Source: Spring 2009 Program Director Self-Administered Questionnaire, 2010, and 2011 Program Director Interviews.

Note: Program director self ratings were collected in 2009 and survey items administered in the program director interview were rated by staff in 2010 and 2011.

[a]In 2009 scores were rated from 1 (low) to 5 (enhanced) while in 2010 and 2011 scores were rated from 1 (low) to 4 (enhanced).

n.a.= not applicable.

Child Development. More than three-quarters of programs across all years (78 to 85 percent) are fully implementing the child development cornerstone. The element of the child development cornerstone that is least likely to be fully implemented, on average, is group socializations (22 percent).

Family Development. Between 66 and 69 percent of Early Head Start programs fully implemented the family development cornerstone in 2010 and 2011. A higher percentage of programs (85 percent) fully implemented the family development cornerstone in 2009, based on directors' self ratings. Parent involvement in program activities is the element that is least likely to be fully implemented in this cornerstone.

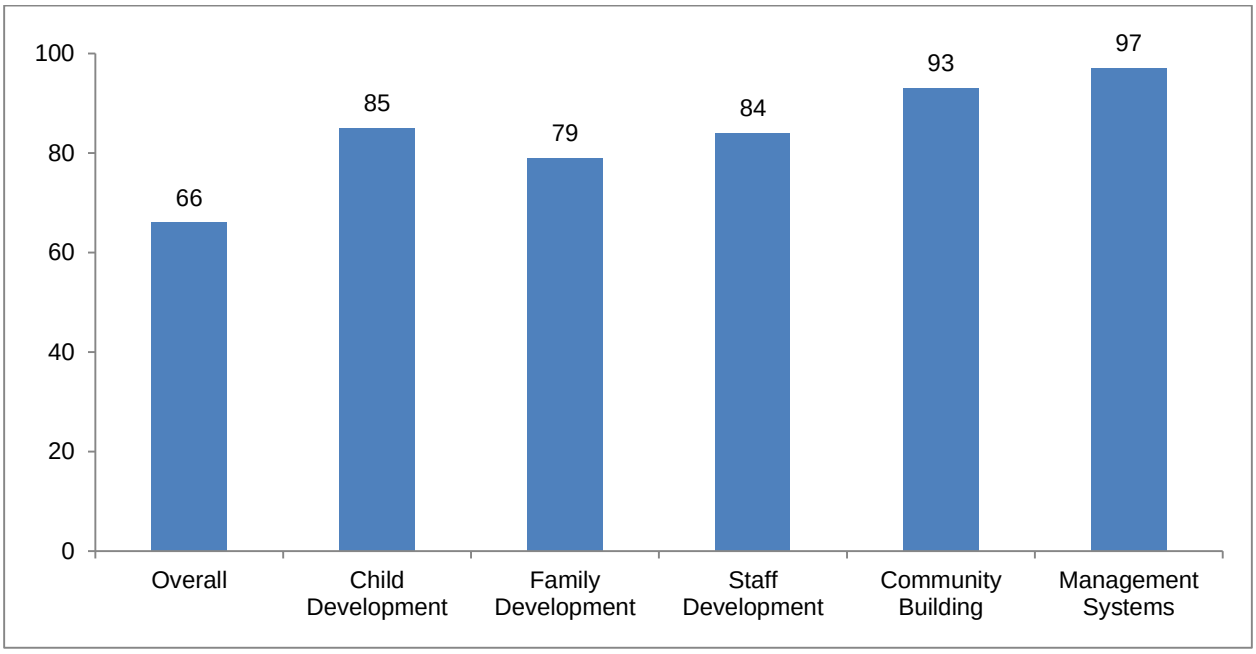
Staff Development. Most Early Head Start programs (82 to 85 percent) fully implement the staff development cornerstone across all years. Self ratings were similar to the ratings derived from the surveys. The element of the staff development cornerstone that is least likely to be fully implemented is turnover (74 percent).

Aggregating Implementation Over Time

Baby FACES has explored various measures of program implementation in an attempt to capture this construct and use it in more complex analyses. In Chapter VII, we examine relationships between program implementation and family participation. For these analyses, we were interested in creating a single variable that captures a program's implementation quality. We sought a measure that capitalizes on the multiple waves of data available but provides a stable indicator of implementation quality that is less prone to year-to-year shifts in program resources, staff turnover, as well as changes in data collection approaches. In keeping with this goal, we created an aggregate measure of implementation that uses data from all three time points. To create this rating, we took the average of ratings for each

cornerstone from all three years to create five cornerstone ratings aggregated over time. Programs with ratings equal to or above 3 on each of the implementation cornerstone ratings are designated as "fully implemented (Appendix D explains the approach for creating an aggregate rating of implementation). Using this definition 66 percent of programs were fully implemented over all three years, with Family Development the least consistently implemented cornerstone, and Management Systems the most consistently implemented (see Figure V.5).

Figure V.5. Percentage Fully Implemented, All Years



Source: Spring 2009 Program Director Self-Administered Questionnaire, 2010, and 2011 Program Director Interviews.

Note: Program director self ratings were collected in 2009 and survey items administered in the program director interview were rated by staff in 2010 and 2011. Programs must have been rated as fully implemented at each time point and on each cornerstone to receive an overall fully implemented rating.

Sample size: N = 89 programs in 2010 and 2011. Sample size in 2009 was 86, and missing data are imputed in the figure.

Summary of Key Findings

- Programs have moderately low staff turnover rates, both for frontline staff (10 and 15 percent of home visitors and teachers, respectively), and at the management level (28 percent lost a manager and 15 percent lost a director).
  - One-third of programs have unfilled full-time staff positions, with an average of two open positions.

- Children have diverse teachers and home visitors.
  - Many children have a teacher (31 percent) or home visitor (34 percent) who speaks a language in addition to English, most often Spanish. English and Spanish are the languages most commonly spoken in classrooms and home visits.
- Most children receive services in their home language.
  - Seventy-eight percent of children from Spanish-speaking homes who receive home-based services have a home visitor who speaks Spanish.
  - For 82 percent of children from Spanish-speaking homes who receive center-based services, a teacher or another adult in the classroom speaks Spanish.
- More than three-quarters of children have a teacher or home visitor with a college degree and experience working with infants and toddlers.
  - About 75 percent of children have a teacher with an AA degree or higher and about 85 percent of children have a home visitor with an AA or higher.
- Children's teachers and home visitors participate in a number of professional development activities, and programs provide staff training opportunities.
  - Approximately one-third of children (31 percent) have a home visitor, and nearly half (45 percent) have a teacher who is currently enrolled in child-care-related training.
  - Children's home visitors report participating in approximately 62 hours of staff trainings annually, while their teachers report participating in about 64 hours of trainings.
- Children's teachers and home visitors report positive feelings about their current jobs and few report depressive symptoms.
- Children and families receiving home visits participate in a variety of activities during visits, with the largest proportion of time spent on child-focused activities (57 percent).
- Most children and families in the home-based option receive home visits in the mid-range of quality, lasting 80 minutes on average.
  - Families have home visits scoring in the mid-range on the HOVRS-A. On the Visitor Strategies subscale, home visits were in the 3-point range (the average subscale score was 3.2). On the Visitor Effectiveness subscale, which captures the home visitor's effectiveness in involving and engaging the family, the average subscale score is 3.8.
- Most children in center-based programs are in classrooms of mid-range quality, with group sizes and ratios within the performance standards.
  - Virtually all children are in classrooms that score a 3 or higher in Emotional and Behavioral Support, with 20 percent in classrooms rated as 6 or higher. In contrast, for Engaged Support for Learning, about half (47 percent) of all children are in classrooms scoring in the low range, with only 2 percent of children in classrooms rated as 6 or higher.
  - Scores are highest in the area of Emotional and Behavioral Support (5.3). Scores are lowest in the area of Engaged Support for Learning (3.3).

- Parents and staff report positive relationships with one another and endorse positive statements about their relationships at similar rates.
- Nearly half of programs are fully implemented overall based on survey ratings.
  - The manner of measuring implementation is important. Director self ratings resulted in more programs reaching full implementation than objective ratings. Self ratings by directors had 72 percent of programs fully implemented compared to objective ratings of survey items of 41 and 45 percent in 2010 and 2011, respectively.
- Overall ratings mask variability in implementation of some cornerstones.
  - The child, and family, cornerstones were implemented less consistently than the staff development and community building cornerstone and the management systems area. This was true for all modes of implementation ratings.
- Two-thirds of programs are fully implemented over all three years.
  - The cornerstone that was least consistently implemented over time was family development; management systems was consistently the highest rated.

VI: QUALITY IN EARLY HEAD START: STABILITY, VARIABILITY, AND PREDICTORS

In Chapter V, we described characteristics of Early Head Start programs and staff and the quality of home visits and classrooms experienced by enrolled children and families. In this chapter, we capitalize on the availability of data at multiple time points to address more complex questions about whether the home visit or classroom quality provided by a particular home visitor or teacher changes over time and whether particular staff or program characteristics are associated with quality.

We fit hierarchical linear models (HLM) to examine variation within staff over time, and between staff within the same program. The HLM models we used allowed us to examine associations between staff characteristics (both time varying—such as depressive symptoms, and time invariant—such as race/ethnicity) and quality. For home visit quality, the HLM models examine: (1) whether home visit quality changes between 2009 and later years, (2) whether home visit quality varies among home visitors within programs, (3) whether home visit quality varies between programs, and (4) whether specific visit content and characteristics and/or home visitor, program, or child/family characteristics contribute to the quality of home visits. Home visit quality was assessed using the Home Visit Rating Scale-Adapted (HOVRS-A; Roggman et al. 2009). For classroom quality, the HLM models examine: (1) whether classroom quality changes between 2010 and later years, (2) whether classroom quality varies between teachers within programs and across programs, and (3) whether teacher and program characteristics are associated with classroom quality. Classroom quality in center-based programs was observed using the Classroom Assessment Scoring System-Toddler (CLASS-T; Pianta et al. 2010a) each year from 2010 to 2012. See Box VI.1 for additional details on the analytic samples and Box VI.2 for additional detail about the models and variables included in each.

Box VI.1. Analytic Samples for HOVRS-A and CLASS-T Models

HOVRS-A: Each home visitor was observed once in each year that she was linked to one or more study children. Over four years, we observed 322 different home visitors at least once. Seven home visitors were observed four times. Seventy-four home visitors were observed three times. Eighty home visitors were observed twice. Finally, 161 home visitors were observed only once. More than half (58 percent) of home visitors observed multiple times were observed with the same family across assessments. The remaining home visitors were observed with two or three different families.

CLASS-T: Each teacher was observed once in the spring each year if she was linked to one or more study children. We have up to three observations for each teacher in the sample (N = 364; 251 were observed once, 106 were observed twice, and 7 were observed three times).

Our analytic sample is not representative of all Early Head Start teachers or home visitors because the Baby FACES study was not designed to collect data on a representative sample of staff. As such, the results of our analyses only pertain to staff who were observed at one or more time points because they were linked to one of our sample children.

Although there are a wide range of possible variables to include in the analytic models given the rich data collected for Baby FACES, we focused on a narrower set of variables as predictors in order to maintain model parsimony. Our approach to selecting predictors for the models was guided by theory and extant research. For example, although evidence for the efficacy of home visiting strategies is mixed (Astuto and Allen 2009; Del Grosso et al. 2011; Paulsell et al. 2010), stronger effectiveness is likely when the quality of the home visit is high and when the relationship between the home visitor and the family is strong. There is also evidence that parents are more engaged and retention is higher when visitors are matched to the family on characteristics such as ethnicity and language (Astuto and

Allen 2009). Classroom quality has been shown to be associated with child–teacher ratio (the number of children per teacher), group size (the total number of children assigned to a teacher or team of teachers), teacher education and training, and teacher salary (Lamb 1998; de Schipper et al. 2006; Vandell and Wolfe 2000). The characteristics of families served also provide important context about classrooms and program functioning and have been linked to Head Start classroom quality (Resnick and Zill 2003). We also solicited input from the Baby FACES TWG. Wherever possible, we selected variables with less missing data and more variability. We sought to minimize redundancy/collinearity, and maintain parallelism with models used in other chapters. The final set of variables was selected to represent key domains and levels (program, staff, family) of interest.

Box VI.2. HLM Models Predicting HOVRS-A and CLASS-T Scores

The level-1 model estimates the variation in observed quality over time. We include dummy variables for each year to capture any differences by year, with year 2009 serving as the reference year in the HOVRS-A models and 2010 as the reference year in the CLASS-T models. Mean quality in the reference year serves as the intercept of the level-1 model.

The level-1 models for the HOVRS-A and CLASS-T both include year-specific depressive symptoms and job satisfaction for each staff member as time-varying covariates (a time-varying characteristic changes from year to year). The level-1 HOVRS-A models also include time-varying visit characteristics and characteristics of the child and family observed: (1) the percentage of time spent on different activities (child-focused, family-focused, and parent-child activities, staff-family relationship building, and crisis management activities), (2) alignment with visit plan, (3) interference from environmental distractions, (4) languages used during the visit, and (5) presence of other children and/or adults during the visit. Some variables are specific to home visit and classroom models. Child and family characteristics, used only in the HOVRS-A analyses, include whether the family enrolled at pregnancy, child age, birth weight, DLL status, race/ethnicity, and maternal demographic risks (all measured at age 1). The level-1 model for the CLASS-T includes time-varying classroom characteristics, that is, staff-parent relationship, adult/child ratio in the classroom, group size, and percentage of DLLs in the classroom at each time point.

Additional time invariant staff characteristics are examined in the level-2 models, including teacher's or home visitor's race/ethnicity; years of experience in Early Head Start; and whether the teacher or home visitor has a bachelor's degree or higher, has a degree in early childhood education, has a CDA credential, ever received support from a mentor or coach, or speaks a language other than English. These time invariant characteristics do not change over time.

The level-3 models include program characteristics. Both the HOVRS-A and CLASS-T models include program approach (center-based versus multiple service options), program implementation status, and indicators of whether the program has a high percentage (more than 50 percent) of families who experience multiple demographic risks, who have mental health or substance abuse problems, or who live in unsafe neighborhoods or experience family violence.

In the tables that follow, the coefficients presented represent average differences in quality associated with each characteristic, holding year of observation and all other characteristics constant. Positive coefficients indicate associations with higher levels of quality and negative coefficients indicate associations with lower quality.

Home Visit Quality

We constructed 3-level HLM models to examine whether home visit quality changed between the initial and later years of observations and whether quality varies between home visitors in the same program and between programs. Results from fitting unconditional models (i.e., models without any predictors) indicated that both Visitor Strategies and Visitor Effectiveness vary significantly between

home visitors and between programs, indicating that the variation in home visit quality is attributable to both staff and program characteristics. To better understand which staff and program characteristics are associated with home visit quality, we included indicators of staff race and ethnicity, years of experience in Early Head Start, whether the home visitor speaks a language other than English, whether she has a bachelor's degree or higher, a degree in early childhood, and/or CDA credential, and whether she ever received support from a mentor or coach. We also include an indicator of whether the home visitor reports any depressive symptoms at each time point. For program characteristics, we include an indicator of whether the program offers multiple service options, an indicator of whether the program is fully implemented, and indicators of whether the program has a high percentage (more than 50 percent) of families who experience multiple demographic risks, who have mental health or substance abuse problems, or who live in unsafe neighborhoods or experience family violence.

Our models also include characteristics of the visit and the child/family observed with the home visitor at each visit. In terms of visit characteristics, we include as predictors the percentage of time spent on different types of activities, the degree of alignment observed between the home visitor's lesson plan and the actual activities that took place, the amount of environmental distractions observed, whether other children or adults were present, and whether other languages were used during the visit. To examine characteristics of the child and family, we include child race and ethnicity, age, whether the family enrolled during pregnancy, and levels of demographic and psychological risks that the family faced at the beginning of the study.

Average Home Visit Quality Does Not Change in Later Years

Descriptive statistics indicate minimal differences in mean home visit quality ratings between year 1 (2009 was the first year of observations) and later years (Table VI.1). The same is true for Visitor Effectiveness ratings. Results from fitting the HLM model confirm that the differences between average home visit quality in year 1 and later years are not statistically significant.

Table VI.1. Mean HOVRS-A Subscale Ratings, by Year

	2009			2010			2011			2012		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Visitor Strategies	232	3.2	0.97	185	3.3	0.82	135	3.3	0.87	19	3.4	0.99
Visitor Effectiveness	231	3.6	0.97	185	3.6	0.90	135	3.7	0.83	19	3.9	0.99

Source: Spring 2009, 2010, 2011, 2012 home visit observations.

Note: Home visit observations were conducted with staff of both the 1-Year-Old Cohort and Newborn Cohort children. Scores range from 1-5.

HOVRS-A= Home Visit Rating Scale-Adapted; SD = Standard deviation.

Home Visit Content and Characteristics are Related to Quality

The percentage of visit time spent on different activities is related to home visit quality (Table VI.2). Compared to a visit spent entirely on child-focused activities, the amount of time spent on staff-family relationship building is negatively associated with Visitor Strategies scores.[50] That is, holding other characteristics constant, visits where more time is spent on staff-family relationship building tend to be of lower quality. In contrast, spending more time on parent-child activities is positively associated with Visitor Strategies. Visitor Effectiveness scores tend to be lower for visits where more time is spent on family-focused activities and staff-family relationship building, and higher for visits where more time is devoted to parent-child activities.[51] The observed degree of alignment of the actual home visit with the home visitor's lesson plan is associated with higher Visitor Strategies and Visitor Effectiveness ratings. The presence of at least one other adult (in addition to the child's parent) during the home visit is associated with higher Visitor Strategies and Effectiveness ratings. The presence of at least one other child during the home visit is associated with lower Visitor Effectiveness. Languages spoken during the visit is not related to quality.

Staff Job Satisfaction Is Related to Quality but Other Staff Characteristics Are Not

Home visitors who are more satisfied with their jobs tend to have higher Visitor Strategies scores. We did not find evidence of differences in home visit quality by staff race and ethnicity. We also found similar levels of quality among home visitors with varying levels of education and experience. Depressive symptoms or experience working with a mentor or coach are not associated with Visitor Strategies or Visitor Effectiveness ratings.

Few Child and Family Characteristics Are Related to Quality

We found that visits conducted with younger children receive lower quality ratings on both Visitor Strategies and Visitor Effectiveness. A one-month difference in child age is associated with a .02 difference in quality ratings. There was a trend to suggest that visits conducted with children who were born with low or very low birth weight receive marginally higher Visitor Strategies ratings, on average. However, child DLL status and demographic and psychological risk, among other characteristics, are not related to quality.

Program Approach and Implementation Are Not Related to Quality

We did not find significant differences in home visit quality ratings between home visitors in programs offering both center- and home-based service options and home visitors in programs offering only a single approach. Home visit quality did not vary by characteristics of the population served by the program. We also found similar levels of home visit quality among programs that are fully implemented and those that are not.

[50] Visitor Strategies, as described in Chapter V, are concerned with the skill exhibited by the home visitor in establishing a relationship with the family, being responsive, facilitating parent-child interaction, and being non-intrusive.

[51] Visitor Effectiveness, as described in more detail in Chapter V, is concerned with parent-child interaction, parent engagement, and child engagement in the visit.

Table VI.2. Home Visit, Child, Family, Staff and Program Predictors of Home Visit Quality

	Visitor Strategies		Visitor Effectiveness	
	Coefficient	Standard Error	Coefficient	Standard Error
Intercept (Mean Quality in 2009)	1.86***	0.32	2.45***	0.32
Home Visit Characteristics				
Percentage of time spent on family-focused activities	0.00	0.00	-0.01**	0.00
Percentage of time spent on parent-child activities	0.01***	0.00	0.01***	0.00
Percentage of time spent on staff-family relationship building	-0.01*	0.00	-0.01*	0.00
Percentage of time spent on crisis management activities	0.01	0.00	0.00	0.00
Alignment with visit plan[a]	0.15***	0.04	0.22***	0.04
Less Interference from environmental distractions[b]	-0.01	0.02	-0.03	0.02
Other children present	-0.09	0.07	-0.15*	0.07
Other adult present	0.20*	0.09	0.20*	0.09
Conducted in another language (vs. English only)	-0.06	0.14	-0.24	0.15
Conducted in multiple languages (vs. English only)	0.23	0.18	0.03	0.18
Child and Family Characteristics				
Enrolled at pregnancy	0.01	0.08	-0.01	0.08
Child age in months	-0.02	0.01	-0.02*	0.01
Child born with Low or Very Low Birth Weight	0.23	0.12	0.10	0.14
DLL	0.07	0.11	0.13	0.12
Race/Ethnicity (vs. white)				
African American	-0.04	0.18	-0.27	0.18
Hispanic	0.15	0.13	0.16	0.13
Other race	-0.07	0.13	-0.07	0.14
Moderate maternal demographic risks[c]	0.16	0.09	-0.02	0.09
High maternal demographic risks[c]	0.04	0.1	-0.10	0.10
One psychological risk factor[d]	-0.04	0.09	0.09	0.09
Two or more psychological risk factors[d]	-0.02	0.13	0.06	0.14
Staff Time-Varying Characteristics				
Staff job satisfaction	0.32**	0.1	-0.04	0.11
Has any depressive symptoms	0.00	0.01	0.01	0.01
Staff Characteristics				
Race/Ethnicity (vs. white)				
African American	-0.24	0.16	-0.15	0.18
Hispanic	-0.08	0.12	0.09	0.14
Other race	0.08	0.19	0.30	0.20
Speaks language other than English	0.02	0.10	-0.08	0.11
Has a BA degree or higher	0.08	0.09	-0.01	0.09
Years of experience in Early Head Start	0.01	0.01	0.00	0.01
Has a degree in early childhood	0.1	0.08	0.11	0.09
Has a CDA credential	-0.01	0.09	0.04	0.09
Ever assigned a mentor or coach	0.10	0.08	-0.02	0.08
Program Characteristics				
Multiple approach[e]	0.15	0.13	0.16	0.13
Population Served				
Over 50% of families with more than 3 demographic risks	-0.01	0.13	-0.20	0.13
Over 50% of families with mental health or substance abuse problems	-0.08	0.17	0.11	0.16
Over 50% of families in unsafe neighborhoods or experiencing family violence	0.08	0.11	0.16	0.11
Fully implemented[f]	-0.10	0.11	-0.09	0.11
Year Effects				
Year 2010	0.04	0.08	-0.02	0.08
Year 2011	0.07	0.09	0.05	0.09
Year 2012	-0.02	0.27	0.19	0.27
Sample Size (Home visitors)	**322**		**321**	

Source: Spring 2009, 2010, 2011 and 2012 Home Visit Observations; Parent Interview; Program Director Interview.

Note: Estimated using a three-level model with random effects for program and staff. Coefficients indicate average differences in home visit rating associated with each characteristic, holding year and all other characteristics constant. For example, a one-percentage point difference in amount of time spent on parent-child activities is associated with a 0.01 difference in Visitor Strategies ratings, holding year and all other characteristics constant.

[a]Assessed on a scale of 1 (not well aligned) to 5 (very well aligned).
[b]Original scale recoded so that 1 (not interfering) to 5 (very interfering). A score of 0 indicates there were no distractions during the visit.
[c]The maternal demographic risk is constructed by summing the number of the following risk factors that the mother faced: (1) being a teenage mother, (2) having no high school credential, (3) receiving public assistance, (4) not being employed or in school or training, and (5) being a single mother.
[d]Family psychological risk is based on moderate or severe depressive symptoms, high parenting stress, and current or past substance use problems.
[e]Programs offering both center- and home-based services (as opposed to only one of the two) are designated as multiple approach.
[f]Programs with ratings equal to or above 3 on each of the implementation cornerstones are designated as "fully implemented".

+$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$.

CDA = Child Development Associate; DLL=Dual language learner.

Classroom Quality

We constructed 3-level hierarchical linear models (HLM) to examine whether classroom quality changed over time and the predictors of classroom quality. Specifically, we examined the domains of Emotional and Behavioral Support and Engaged Support for Learning of the CLASS-T. The results indicate that both domains of the CLASS-T vary significantly between teacher/classrooms and between programs, suggesting that the variation in classroom quality is attributable to both staff/classroom characteristics and program characteristics.

Engaged Support for Learning Declines Over Time, While Emotional and Behavioral Support Remains Unchanged

Examining change over time in classroom quality descriptively shows that the mean for Engaged Support for Learning decreases by about a third of a point between 2010 and 2012 [3.56 in 2010 (SD = 1.21) versus 3.25 in 2012 (SD = 0.67) (Table VI.3)]. Results from multilevel models confirmed this trend (Table VI.4).

In contrast, the mean for Emotional and Behavioral Support remains consistent over time. The means range from 5.23 to 5.30 (Table VI.3). Results from multilevel models indicate no significant differences in Emotional and Behavior Support across years (Table VI.4).

Teacher Educational Level Is Significantly Associated with Engaged Support for Learning

Among the staff characteristics examined in the models, having a BA or higher is significantly associated with higher scores in Engaged Support for Learning; however, it is not associated with Emotional and Behavioral Support. None of the other staff characteristics in the model are associated with classroom quality.

Table VI.3. Descriptive Statistics for the CLASS-T Domains Across Time

	2010			2011			2012		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Engaged Support for Learning	218	3.56	1.21	226	3.19	1.24	40	3.25	0.67
Emotional and Behavioral Support	218	5.29	0.83	226	5.23	0.90	40	5.30	0.59

Source: Spring 2010 to spring 2012 classroom observation.

Note: Classroom observations were conducted with staff of both the 1-year-old Cohort and Newborn Cohort children. Scores range from 1-7.

CLASS-T=Classroom Assessment Scoring System-Toddler; SD = Standard deviation.

Table VI.4. Staff, Classroom, and Program Predictors of Classroom Quality

	Engaged Support for Learning		Emotional and Behavioral Support	
	Coefficient	Standard Error	Coefficient	Standard Error
Intercept (Mean Classroom Quality in 2010)	3.21***	0.33	4.93***	0.22
Staff Characteristics				
Race/Ethnicity (vs. white)				
African American	-0.08	0.19	-0.06	.013
Hispanic	-0.16	0.21	0.04	0.12
Other race	-0.09	0.26	-0.09	0.18
Speaks language other than English	-0.08	0.16	0.13	0.09
Years of experience in Early Head Start	0.01	0.02	0.01	0.01
Has a BA degree or higher	0.27*	0.13	0.10	0.09
Has a degree in early childhood	0.05	0.12	0.02	0.08
Has a CDA credential	0.01	0.11	0.08	0.07
Ever assigned a mentor or coach	-0.08	0.14	-0.04	0.07
Staff/Classroom Time-Varying Characteristics				
Staff depressive symptoms	-0.01	0.01	-0.03*	0.01
Staff-parent relationship[a]	0.16*	0.07	0.12**	0.05
Staff job satisfaction	0.31*	0.15	0.24*	0.12
Adult/child ratio	0.10	0.08	-0.01	0.05
Class size	-0.04	0.04	-0.03	0.03
Percentage DLLs in the classroom	-0.00	0.00	0.003*	0.001
Program Characteristics				
Multiple approach[b] (versus single approach)	-0.06	0.22	0.06	0.12
Population served				
Over 50% of families with more than 3 demographic risks	0.07	0.21	0.15	0.13
Over 50% of families with mental health or substance abuse problems	0.21	0.21	0.09	0.16
Over 50% of families in unsafe neighborhoods or experiencing family violence	-0.03	0.16	-0.17	0.12
Fully implemented	0.14	0.17	0.04	0.11
Year Effects				
Year 2011	-0.34**	0.13	-0.05	0.08
Year 2012	-0.32*	0.15	0.02	0.13
Sample Size (Teachers)	**364**			

Source: Spring 2010 to spring 2012 classroom observation, Staff Interview, and Program Director Interview.
Note: Classroom observations were conducted with staff of both the 1-year-old Cohort and Newborn Cohort children.
[a]This represents the average staff-reported staff-parent relationship calculated across sample children in the classroom.
[b]Programs offering both center- and home-based services (as opposed to only one of the two) are designated as multiple approach.

*p < .05; **p < .01; ***p< .001.

CDA = Child Development Associate; DLL=Dual language learner.

Teacher Job Satisfaction and Teacher-Parent Relationship Are Positively Associated with Classroom Quality Over Time

Teacher job satisfaction and teacher-parent relationship[52] are significantly positively associated with higher scores over time in both Emotional and Behavioral Support and Engaged Support for Learning. Staff members who reported that they were very likely to continue working in early childhood education have higher scores in both Emotional and Behavioral Support and Engaged Support for Learning than staff who do not. As rated by staff, the teacher-parent relationship is significantly and positively associated with higher scores in both Emotional and Behavioral Support

[52] Reported by staff.

and Engaged Support for Learning. That is, the higher the levels of teacher-parent relationships, the higher the scores in Emotional and Behavioral Support and Engaged Support for Learning.

Teacher Depressive Symptoms and Percentage of DLLs in the Classroom Are Associated with Emotional and Behavioral Support Over Time

Among time-varying teacher/classroom characteristics, greater depressive symptoms in teachers are significantly associated with lower scores in Emotional and Behavior Support. Classrooms with greater concentrations of DLLs have higher scores in Emotional and Behavioral Support. Neither variable, however, predict Engaged Support for Learning. Adult/child ratio and class size are not associated with either domain of classroom quality.

Program Characteristics Are Not Associated with Classroom Quality

We did not find any of the select program characteristics included in the models to predict classroom quality. Other program characteristics that we did not collect (such as the quality of teacher professional development) may explain the variation in classroom quality among programs.

Summary of Key Findings

- Home visit quality:
  - Differences in mean home visit quality ratings between the first year of data collection and later years are small and not statistically significant.
  - Home visit content and characteristics are related to home visit quality.
    - The amount of time spent on staff-family relationship building is negatively associated with Visitor Strategies scores.
    - Visitor Effectiveness scores tend to be lower for visits where more time is spent on family-focused activities and higher for visits where more time is devoted to parent-child activities.
    - The observed degree of alignment of the actual home visit with the home visitor's lesson plan is associated with higher Visitor Strategies and Visitor Effectiveness ratings.
    - The presence of at least one other adult during the home visit is associated with higher Visitor Strategies and Effectiveness ratings. The presence of at least one other child during the home visit is associated with lower Visitor Effectiveness ratings.
  - We did not find evidence of differences in home visit quality by staff race and ethnicity, education, experience, or mental health status.
  - Few child and family characteristics are related to home visit quality.
    - Visits conducted with younger children received lower quality ratings on both Visitor Strategies and Visitor Effectiveness.
    - Child DLL status and demographic and psychological risk, among other characteristics, were not related to quality.
  - Program approach and implementation are not related to home visit quality.

- Classroom quality:
  - Engaged Support for Learning declines over time, while Emotional and Behavioral Support remains unchanged.
  - Teacher educational level is significantly associated with Engaged Support for Learning.
  - Teacher job satisfaction and teacher-parent relationship are positively associated with classroom quality over time.
  - Teacher depressive symptoms are negatively associated and percentage of DLLs in the classroom are positively associated with Emotional and Behavioral Support over time.
  - None of the program characteristics in the models are associated with classroom quality.

VII: EARLY HEAD START SERVICES: PROGRAM OFFERINGS AND FAMILY PARTICIPATION

The potential benefits that families can reap from Early Head Start depend on the extent of their participation in program activities; the expectation is that higher or more intense participation yields the most benefit (Bagnato et al. 2011; Berkel et al. 2010). Berkel et al. (2011) posits that program effects are moderated by participant responsiveness, which includes not only satisfaction and engagement but "enthusiastic participation," and dosage.[53] Quantifying program participation is important and is captured in a number of ways in Baby FACES. One way we characterize participation is by describing the length of time a family remains enrolled in the program. Because Early Head Start serves pregnant women and children between the ages of 0 to 3 years, families who enroll while the mother is pregnant have a longer period of eligibility during which they can receive services. Families who enter later or leave the program before the child turns 3 are in this sense participating at a lower level than those who remain in the program longer and it is possible that these families do not receive the maximum benefits of program services. A second way we measure participation is by assessing whether families are actively involved in program activities while they are enrolled. A third way we measure participation is by quantifying the amount of specific services families receive. OHS provides guidance pertaining to the services that programs should *offer* to families, but recognizes that families' choices influence the services they receive. While families' participation is voluntary, it is recommended that programs make every effort to encourage and support high rates of home visit completion and center attendance.

In this chapter, we use information from program directors and staff to describe program offerings and family participation. We begin by documenting how long families stayed enrolled in the program. Next, we examine whether families demonstrate active and consistent involvement in program activities based on annual staff ratings. Finally, we describe the amount of services programs offered and the amount children and families received during their enrollment based on weekly reports collected from program staff (the FST system).[54] We also determine the extent to which programs offer the amount of services recommended by OHS. Until now, there have been no studies of Early Head Start that have collected ongoing weekly information about the services that families and children receive throughout their time in the program. Although our measures are not complete records of all services offered by programs or received by families, they move us much farther down the path to understanding what enrollment in Early Head Start means.[55]

[53] Berkel et al. (2011) use the general term "responsiveness" to describe various aspects of program participation, such as participant satisfaction, and dosage or service receipt.

[54] Program staff reported weekly on the services that were offered and received by each study family, including reasons for missed services. See Chapter II for a detailed description of the Family Services Tracking system.

[55] We were able to collect information on home visits offered by programs and completed by families and center days offered by programs and attended by families as reported by staff. We were unable to collect information on screenings and assessments, referrals made based on screening/assessment or other family needs, or attendance at parent education, and therefore recognize that we gathered just one aspect of program services. (Appendix D of the Baby FACES Age 2 Report [Vogel et al. 2015] describes iterations in administration of the FST.)

Box VII.1. Measures of Participation

Length of Program Enrollment. This is a measure of the number of months elapsed between a child's enrollment and exit dates or the end of Baby FACES data collection when children were 3 (whichever came first).[56] Dates of enrollment were provided by programs and exit dates were reported by programs through the FST system and confirmed by Mathematica staff. We also obtained exit date information when contacting programs for annual site visits,. We are not able to observe exit dates for those who were still enrolled at age 3; thus for these analyses we chose March 1, 2011 as the end date for children in the 1-year old Cohort and March 1, 2012 for children in the Newborn Cohort. Fifty-nine percent of the 1-year-old Cohort and 45 percent of the Newborn Cohort had not exited or aged-out of the program prior to these dates. Using the same end date for each cohort provides a uniform study period for all children.

Family Involvement. Teachers or home visitors complete one item to rate each family's level of involvement in the program over the past six months. This is collected on up to three occasions for each family during annual data collection visits. This same measure of involvement was used in the EHSREP. Ratings were assigned on a 4-point scale:

- 4 – Family was consistently highly involved in the program over the past six months—the family kept most appointments, actively engaged in home visits and group activities, and child attended center regularly (for families in the center-based option only).
- 3 – Family's involvement in the program varied during the past six months—the family was sometimes highly involved in the program, and at other times the family's involvement was low.
- 2 – Family's involvement in the program was consistently low throughout the past six months—the family kept some appointments but missed or canceled frequently, family did not actively engage in home visits and group activities, and child was often absent from the center (for families in the center-based option only).
- 1 - Family was not involved in the program at all.

Services Offered and Received. This denotes the number of home visits and center days offered by programs and received by families over a two year period. Service provision and family participation were tracked on a weekly basis by program staff using the FST when children were between 1 and 3 years of age and enrolled in the program. We do not calculate these figures for children in the combination service option (less than one percent of the sample). For the Newborn Cohort, data were collected from July 2010 to June 2012. For the 1-year-old Cohort, data were collected from July 2009 to June 2011.

- **Amount of Services Offered** – This denotes the total number of home visits or center days that were available for families to receive while they were enrolled and between the 1 and 3 years if age. We count the number of home visits or center days scheduled per week and exclude services that were missed due to program closures or staff illness.
- **Amount of Services Received** – This denotes the total number of home visits or center days received by families during their enrollment. We count the number of home visits or center days families participated in and exclude those missed due to either family reasons (such as cancelling a visit or not showing up), or program reasons (such as program closure, or staff illness).
- **Services Offered and Received per Week** – This denotes the weekly average home visits or center days offered and received. We calculate the average in each year by dividing services offered or received by the number of weeks that a family spent in the program in that year. The weekly average for families who are enrolled for a full year is based on 52 weeks. Families who leave the program before the end of the FST data collection period in each year have weekly averages based on fewer weeks.
- **Service Take-Up Rate (Percentage of Home Visits/Center Days Offered that Were Received)** – Take-up rate is calculated by dividing the total number of home visits or center days completed or attended by the total number of home visits or center days offered.

[56] During spring 2009, we recruited and enrolled children who were study eligible because they fell into one of our birthday windows (8 weeks or younger for the Newborn Cohort, and between 10 to 15 months for the 1-year-old Cohort). Therefore, children in both cohorts had been enrolled in the program for varying periods of time (potentially longer than one year if a 1-year-old Cohort child enrolled prenatally).

Most Families Attend the Program for more than Two Years but Families Enroll and Exit at Different Periods

Among children who enrolled in Early Head Start at age 1 or earlier, the average length of enrollment is 28 months (Table VII.1). However, the average masks considerable variability in this measure, which ranges from 0.5 to 43 months. Several factors may account for some of the variability in length of program enrollment, namely when families enroll and when families exit.

Although Early Head Start enrolls pregnant women, only some families enrolled before the child was born. Thirty-five percent of families enrolled while the mother was pregnant (82 percent of the Newborn Cohort and 24 percent of the 1-year-old Cohort). Provided these families stay until their eligibility ends, they would presumably have more exposure to the program. Data from Baby FACES indicate that among families who enrolled during pregnancy, the average length of enrollment is 33 months. In contrast, the average length of enrollment among children whose families enrolled after birth (and by 15 months of age) is lower, 25 months.

Table VII.1. Length of Program Enrollment in Months

	Mean (SE)	Range	Sample Size
All Families	27.9 (0.44)	0.5-44.0	**971**
Families Who Enrolled During Pregnancy	33.0 (0.72)	2.6-44.0	**344**
Families Who Enrolled Postpartum	25.1 (0.41)	0.5-37.9	**569**
Families Who Were Early Leavers (Left Early Head Start before age 3)	16.6 (0.57)	0.5-37.7	**361**

Source: Sample Management System (SMS) through March 1, 2012.

Note: Length of enrollment is calculated as the number of months elapsed between a child's enrollment date and exit date or the end of data collection (March 1, 2011 or 2012 for the Newborn and 1-year-old Cohort, respectively).

SE = Standard error.

Program exits are a second potential source of variability in the length of enrollment. Children are eligible to participate in Early Head Start until age 3 when they can transition or "graduate" to Head Start or another early childhood program serving preschool children. However, some children leave the program before their eligibility has ended.[57] In our sample, 37 percent of children left the program before they were 3 years old (56 percent of the Newborn Cohort and 33 percent of the 1-year-old Cohort). The average length of enrollment for early leavers is 17 months (ranging from 1 to 38 months). Average child age at exit for early leavers is 17 months and ranges from 0 to 31 months. See Caronongan et al. (2014) for additional information on those who exit the program early.

Many Families Demonstrated High Family Involvement Particularly at Earlier Ages

The majority of families who remained enrolled at each data collection point were rated by their teacher or home visitor as consistently highly involved at ages 1 and 2. This is true whether we examine involvement ratings for the full sample of families or restrict the sample to families who do not leave their program early. At age 1, 60 percent of families in the full sample are rated as highly involved (Table VII.2). The percentage is slightly higher (65 percent) when we look only at families who did not eventually leave their programs early. At age 2, the percentage of families rated as highly involved is nearly identical (56 percent) whether we look at the full sample or restrict to families who remain

[57] In exit interviews families reported on the most common reason for leaving: the family moved away (37 percent, SE=5.65); the family wanted their children in the same program (29 percent, SE=5.00); the center or home visit hours were inconvenient (20 percent, SE 3.71); and the family was too busy to participate (19 percent, SE= 3.09).

enrolled through age 3. By age 3, among families still enrolled, the percentage of highly involved families declines to 49 percent. Among 405 families with complete staff ratings across 3 years, 27 percent (SE=2.81) are rated as highly involved at all time points (not shown).

Table VII.2. Staff Ratings of Family Involvement, by Child Age

	Percentage (SE)		
	Age 1	Age 2	Age 3
Full Sample			
Family Involvement in the Past Six Months			
Consistently high involvement	59.5 (2.41)	55.5 (2.92)	48.5 (2.74)
Varied involvement	32.0 (1.98)	35.4 (2.56)	42.6 (2.80)
Consistently low involvement	7.2 (1.0)	8.4 (1.35)	7.6 (1.35)
Not involved	1.3 (0.68)	0.7 (0.40)	1.2 (0.64)
Sample Size	**737**	**589**	**497**
Families Who Remain Enrolled Until Age 3			
Family Involvement in the Past Six Months			
Consistently high involvement	65.1 (2.61)	55.5 (2.78)	48.5 (2.74)
Varied involvement	27.3 (2.16)	36.9 (2.63)	42.6 (2.80)
Consistently low involvement	6.7 (1.24)	6.7 (1.27)	7.6 (1.35)
Not involved	0.9 (0.54)	0.8 (0.46)	1.2 (0.64)
Sample Size	**518**	**519**	**497**

Source: Staff-Child Report (SCR) Spring 2009, 2010, 2011, 2012.

Note: For children in the home-based option, the SCR was filled out by the family's home visitor. For children in the center-based option, the SCR was filled out by the child's teacher. For the few children receiving the combination option, either the teacher or home visitor was randomly selected to complete the reports (and that same category of staff remained responsible for reports unless the child changed to a different service option).

SE=Standard error.

Very few families (7 to 8 percent) are reported to show consistently low involvement in the three years observed, whether or not we exclude early leavers from the analytic sample. Even fewer families (approximately 1 percent) are rated as not involved at any age. At age 3, even after some families had left their programs, the percentage of families who are not involved remained at 1 percent.

Programs Offered Home Visits and Center Days at High Frequencies

The average number of home visits offered per week to families in the home-based option was 0.92 when children were between the ages of 1 and 2 and 0.93 between the ages of 2 and 3 (Table VII.3). The average number of center days offered per week was greater than 4 in both years. The OHS recommends that home-based families are offered at least 48 home visits per year.

Table VII.3. Services Offered by Programs and Received by Families[a]

	Mean (SE) Ages 1-2[b]	Mean (SE) Ages 2-3[b]
Services Offered per Week		
Number of home visits offered per week	0.92 (0.01)	0.93 (0.01)
Number of center days offered per week	4.3 (0.07)	4.2 (0.07)
Total Services Offered per Year		
Total home visits offered	47.6 (0.67)	43.7 (0.78)
Total center days offered	224.5 (3.84)	193.8 (4.65)
Percentage of children offered at least 48 visits during the year[c]	59.5 %	40.9 %
Percentage of children offered at least 240 center days during the year[c]	36.7 %	18.2 %
Services Received per Week		
Number of home visits completed per week	0.70 (0.02)	0.69 (0.02)
Number of center days attended per week	3.6 (0.07)	3.7 (0.08)
Total Services Received per Year		
Total home visits completed	36.1 (0.82)	32.6 (0.92)
Total center days attended	189.4 (3.85)	167.0 (4.53)
Percentage of children who completed at least 48 visits during the year	9.4 %	6.7%
Percentage of children who attended at least 240 center days during the year	11.0 %	6.1 %
Service Take-up Rates (percentage)		
Home visit take-up rate[d]	77.3 (6.77)	74.2 (1.98)
Center days take-up rate[d]	85.1 (4.80)	85.4 (8.26)
Sample Size	**247-254**	**214-256**

Source: Family Service Tracking (FST) through March 1, 2011 for the Newborn Cohort and through March 1, 2012 for the 1-year-old Cohort.

Note: The weekly average for families who are enrolled for a full year is based on 52 weeks. Families who leave the program before the end of the FST data collection period in each year have weekly averages based on fewer weeks.

[a]Estimates calculated for families who did not change service options or exit the program early using multiply imputed data.
[b]Ages 1-2 includes the services children received between age 1 and 2. Ages 2-3 includes the services children received between ages 2 and 3.
[c]OHS recommends that programs offer at least 48 visits per year and 240 center days.
[d]Take-up rate is the percentage of home visits or center days offered that were completed or attended.

SE=Standard error.

Among families who participated for a full year and did not change service options,[58] the average total number of home visits offered was 48 between age 1 and age 2 and 44 between age 2 and age 3. Sixty percent of children in the home-based option were offered at least 48 home visits between ages 1 and 2. Forty-one percent were offered this number between ages 2 and 3. For children in the center-based option, OHS recommends offering 240 days per year or just under 5 days per week on average. Center-based families in Baby FACES were offered an average of 225 total days between ages 1 and 2 and 194 days between ages 2 and 3. Thirty-seven percent of children in the center-based option were offered the recommended 240 center days between ages 1 and 2. Eighteen percent were offered this number between ages 2 and 3.

[58] Changes in service option were relatively uncommon: of 370 children in the home-based option from ages 1 to 2, just 8 changed to center-based only from ages 2 to 3, and became part of the center-based analytic sample for that year; of 331 children in the center-based option from ages 1 to 2, none changed to the home-based option from ages 2 to 3.

Families Did Not Take Up All the Home Visits or Center Days that Were Offered by Programs

The average number of home visits completed per week was 0.7 in both years (Table VII.3). These weekly averages indicate that between 2 to 3 home visits are completed per month. The average number of center days children attended per week is 3.6 when children were between ages 1 and 2 and 3.7 between ages 2 and 3.

Among families who participated for a full year and did not change service options, the average total number of home visits completed is 36 between ages 1 and 2 and 33 between ages 2 and 3. Children in the center-based option attended a total of 189 center days when between ages 1 and 2 and 167 center days between ages 2 and 3, on average.

Because some of the variability in services received is possibly due to differences in program offerings, we calculated service take-up rates (the percentage of offered services actually received) for families in a given option during the course of a full year. Take-up rates help to clarify the extent to which variability in services received is due to families' decisions and behavior. Consistent with findings about services offered and received in other types of early childhood and family support interventions, the services received by families are lower than the amounts offered by programs, indicating that families did not take up all of the home visits or center days that were available to them. Take-up rates for home visits and center days indicate that families typically participated in 74 to 85 percent of home visits and center services offered. Families in the home-based option for a full year completed 77 percent of the home visits they were offered between ages 1 and 2, and 74 percent between ages 2 and 3. Early exiters had lower home visit take-up rates compared to children who did not exit early while they were enrolled between ages 1 and 2 (60 percent, SE=5.56), but similar take-up rates between ages 2 and 3 (60 percent, SE=16.52; not shown). Children who were in the center-based option for an entire year (between either or both ages 1 to 2 and ages 2 to 3) attended 85 percent of center days offered between ages 1 and 2 and between ages 2 and 3, on average Center-based children who left their programs early attended a similar amount of the center days offered while they were enrolled. Their average center take-up rate was 57 percent (SE=73.24) between ages 1 and 2 and 81 percent (SE=40.05) between ages 2 and 3 (not shown).

Summary of Key Findings

- Most families attend the program for more than two years but families enroll and exit at different periods.
  - Among children who enrolled in Early Head Start at age 1 or earlier, the average length of enrollment is 28 months.
  - Among families who enrolled during pregnancy, the average length of enrollment is 33 months. In contrast, the average length of enrollment among children whose families enrolled after birth (and by 15 months of age) is 25 months.
  - The average length of enrollment for early leavers is 17 months.
- Many families demonstrate high family involvement particularly at earlier ages.
  - A majority of families are rated by staff as consistently highly involved at ages 1 and 2.
  - By age 3, the percentage of families demonstrating high involvement is 49 percent compared to 60 percent and 56 percent at ages 1 and 2, respectively.

- Programs offered home visits and center days at high frequencies.
  - On average, families in the home-based option are offered 0.92—0.93 home visits per week when children are between ages 1 to 3. The average total number of home visits offered in each year ranges from 44 to 48.
  - The average number of center days offered per week to families in the center-based option is greater than 4 in both years. The average total center days offered is 225 between ages 1 and 2 and 194 days between ages 2 and 3.
- Families do not take up all the home visits or center days that are offered by programs.
  - Families in the home-based option for a full year complete 77 percent of the home visits they were offered between ages 1 and 2, and 74 percent between age 2 and 3.
  - Children in the center-based option for a full year attend 85 percent of center days offered between ages 1 and 2 and between ages 2 and 3, on average.
  - Children who eventually left their programs early have lower home visit take-up rates while they are enrolled.

VIII: WHAT PREDICTS PARTICIPATION IN EARLY HEAD START?

Chapter VII defined participation in different ways and described the participation of the average child and family enrolled in Early Head Start. In this chapter, we delve further into how participation varies between programs and families. Specifically, the goals of this chapter are to examine (1) how much participation varies depending on the program a family is enrolled in, (2) what child and family characteristics predict participation, and (3) what staff and program characteristics predict participation.

As described briefly in Chapter II and more fully in Appendix D, we use multilevel models to examine each measure of participation. Multilevel models take into account that we have multiple observations of children over time and that families enrolled in the same program may have similar patterns of participation because of their shared experiences. For example, a program that more actively reschedules missed home visits may have families with a higher rate of home visit receipt than families in a program that does not make up missed visits in a similar manner. If we compare families from different programs without accounting for their program affiliation, we may mistakenly conclude that the differences we find are due to family characteristics when, in fact, at least part of the difference is due to the programs they attend. To examine how much participation varies over time, among children or families, and between programs, we first fit "unconditional" multilevel models (that is, models without any predictors). We then test the relative influence of child, family, staff and program-level characteristics on participation by including them as predictors.

Examining Variation in Participation

We examine how much of the total variability in each of the participation measures is due to differences between programs, among families in the same program, and within families over time by fitting unconditional multilevel models without any predictors. We estimate variability between programs by comparing how much the average participation of each program differs from the average of all programs. To estimate variability between families, we compare how much each family's participation (averaged across multiple observations, if available) differs from the average participation of all families in the same program. Finally, we estimate how participation varies over time by comparing how much participation measured on a particular occasion for a particular family differs from that family's average participation across observations. Families who are observed on only one occasion do not contribute to this estimate.

Participation Is Not Substantially Different Across Programs

In unconditional models (without covariates), we found little evidence to suggest that the particular program a family attends explains their attendance and participation in Early Head Start. Twenty-three percent of the variability in length of enrollment is due to differences between programs. This indicates that the average length of stay is about the same regardless of the particular program that a family attends. Similar to length of enrollment, the probability of high involvement does not differ substantially between programs. Only 7 percent of the variability in involvement is due to program-level differences. Service take-up rates also do not vary considerably by program. Less than a quarter of the variability in home visit completion and center attendance is due to differences between programs.

Participation Varies Substantially Among Families Enrolled in the Same Program

There is much stronger evidence to suggest that characteristics of individual families explain their decisions and choices about participation in the program. Seventy-eight percent of the variability in length of enrollment is due to differences between families. Families attending the same program differ considerably in their levels of involvement, with 52 percent of the variability due to differences between families. Families also differ considerably in terms of their service take-up rates. Sixty-six percent of the variability in home visit take-up rates and 61 percent of the variation in center take-up rates is due to between-family differences.

Family Involvement and Service Take-Up Rates Vary Over Time

Because family involvement and service take-up were assessed multiple times for many families, we were also able to examine how much participation varies within an individual family from one occasion to another. Forty-one percent of the variation in family involvement is due to differences observed within families over time. This indicates that a family rated as highly involved in one year may not necessarily be rated as such in another year. Service take-up rates also vary over time, but to a lesser extent. Just 10 percent of the variability in home visit take-up rates and 16 percent of the variability in center take-up rates is due to year-to-year differences within families. This indicates that families' service take-up rates are fairly stable or consistent from one year to the next.

Predictors of Length of Enrollment

For length of enrollment, we limit our analyses to child and family characteristics that do not change over time. Because most enrollment dates precede the beginning of the Baby FACES study period, we cannot assume that information collected at the beginning of the study about family characteristics that can change over time (such as maternal employment status) reflect their characteristics at initial program enrollment. Similarly, children who are early leavers have data that are non-randomly missing in data collection periods following their exit dates.

We do not include teacher or home visitor characteristics as predictors for length of enrollment because the staff for whom we have data may not be the staff person assigned to the child at enrollment. Staff changes are fairly common; 46 percent (SE = 3.46) of children in the sample experienced at least one staff change over 3 years.

Length of Enrollment Varies by Child and Family Characteristics

We did not find differences in length of enrollment by child race/ethnicity or gender. However, children who are DLLs stay in Early Head Start for almost four months longer (Table VIII.1) compared to children from homes in which English is the only language spoken. We found marginally shorter enrollment among those with medium demographic risk versus lower risk (at a trend level). In terms of family risk factors, children whose mothers face medium demographic risks were enrolled for a shorter period (1.3 months less) compared to children facing lower maternal demographic risks.[59] We did not find a difference in enrollment length between children facing high versus lower maternal demographic risks. Length of enrollment also did not significantly differ by the levels of psychological risks faced by parents.

[59] See Chapter II for a description of the measure of risk.

Table VIII.1. Predictors of Length of Enrollment in Early Head Start

	Coefficient	Standard Error
Child and Family Characteristics		
Race/Ethnicity (vs. white)		
African American	1.27	1.53
Hispanic	-1.80	1.41
Other	-0.32	1.23
Male child	-0.28	0.71
DLL	3.76***	1.07
Child born with low or very low birth weight	-1.02	2.40
Maternal risk[a] (versus lower risk)		
Medium maternal demographic risks	-1.30+	0.75
High maternal demographic risks	-1.70	1.16
Psychological risk[b] (versus zero risks)		
One psychological risk factor	-1.01	1.00
Two or more psychological risk factors	-1.62	1.30
Program Characteristics		
Multiple approach[c] (versus single approach)	-1.77	1.21
Population served		
Over 50% of families with mental health or substance abuse problems	-0.26	1.81
Over 50% of families in unsafe neighborhoods or experiencing family violence	-0.23	1.17
Over 50% of families with more than 3 demographic risks	1.59	1.26
Fully implemented[d]	0.90	1.15
Sample Size	**971**	

Source: Survey Management System (SMS), Family Service Tracking (FST); Staff Child Report; Parent Interview; Program Director Interview.

Note: Estimated using a two-level model with random program effects. Length of enrollment is calculated as the number of months elapsed between a child's enrollment date and exit date or the end of data collection (March 1, 2011 or 2012 for the Newborn and 1-year-old Cohort, respectively). Coefficients indicate differences in length of enrollment (in months) associated with each characteristic. For example, length of enrollment for children in the center-based option was 1.13 months shorter, on average, compared to children in the home-based option.

[a]The maternal demographic risk is constructed by summing the number of the following risk factors that the mother faced: (1) being a teenage mother, (2) having no high school credential, (3) receiving public assistance, (4) not being employed or in school or training, and (5) being a single mother.
[b]Family psychological risk is based on moderate or severe depressive symptoms, high parenting stress, and current or past substance use problems.
[c]Programs offering both center- and home-based services (as opposed to only one of the 2) are designated as multiple approach.
[d]Programs with ratings equal to or above 3 on each of the implementation cornerstones are designated as "fully implemented".

+$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$.

DLL=Dual language learner.

Length of Enrollment Is Not Related to Program Approach, Population Served, or Program Implementation

We did not find significant differences in average length of enrollment among families in programs offering both center- and home-based services (multiple-approach programs) compared to those attending programs offering only a single approach (Table VIII.1). Length of enrollment also did not differ for programs that serve a high proportion of families with psychological or demographic risks. Families in programs that were rated as fully-implemented have enrollment lengths that were similar to families in programs that were not fully-implemented.

Predictors of Family Involvement

To examine predictors of family involvement, we fit models comparing characteristics of families who are rated by staff as consistently highly involved in the program with families who are not rated as highly involved (including families rated as having varying involvement, consistently low, or no involvement). We included as predictors the same set of child, family, and program characteristics that we used to predict length of enrollment. In addition, we also included an age-specific indicator of child health in order to determine whether concurrent child health status is related to family involvement. We also included characteristics of the primary staff member families were working with at each data collection point.

Family Involvement Attenuates Over Time

Compared to involvement at age 1, the probability of being highly involved is 8 percentage points lower at age 2, and 15 percentage points lower at age 3 (Table VIII.2). This finding is consistent with the findings described earlier, in which we found a high percentage of variability in involvement within families over time.

Family Involvement Is Related to Child and Family Characteristics

Families who enrolled during pregnancy (and have therefore been enrolled longer) were not more likely to be highly involved than families who enrolled after the child was born (Table VIII.2). However, families who were early leavers have a lower probability (by 14 percentage points) of being highly involved compared to families who stayed through age 3. We did not find a difference in involvement among families by service option.

Hispanic families are marginally more likely to be highly involved compared to white families although this difference only trended toward significance. We did not find a significant difference in the probability of high involvement of African American families compared to white families. In terms of family risk factors, families facing medium maternal demographic risk have a lower probability of being rated as highly involved (7 percentage points lower) compared to those facing lower maternal demographic risk. Likewise, families facing the highest maternal demographic risk also have a lower probability of a high involvement rating compared to families with lower risk (7 percentage points lower).

Few Staff Characteristics Are Related to Family Involvement

We found limited evidence of associations between staff characteristics and family involvement. Staff education, experience, and language spoken were not associated with family involvement.

Family Involvement Is Not Related to Program Approach, Population Served, or Program Implementation

We did not find significant differences in family involvement among families in programs with varying characteristics. Specifically, families in multiple-approach programs have a similar probability of being highly involved compared to those attending programs offering only a single approach (Table VIII.2). Family involvement is also not associated with characteristics of the program population served or program implementation.

Table VIII.2. Predictors of High Family Involvement

	Coefficient	Standard Error
Age 2	-0.08**	0.03
Age 3	-0.15***	0.03
Child and Family Characteristics		
Family enrolled in pregnancy	0.04	0.03
Family left Early Head Start early (before age 3)	-0.14***	0.03
Service Option (vs. Home-based)		
Center-based	-0.00	0.07
Combination	-0.00	0.42
Race/Ethnicity (vs. white)		
African American	-0.04	0.06
Hispanic	0.10+	0.05
Other	0.03	0.06
Male child	-0.00	0.03
DLL	-0.00	0.04
Child born with low or very low birth weight	0.00	0.06
Child in excellent or very good health	-0.07	0.04
Maternal risk[a] (versus lower risk)		
Moderate maternal demographic risks	-0.07*	0.03
High maternal demographic risks	-0.07*	0.04
Psychological risk[b] (versus zero risks)		
One psychological risk factor[b]	-0.01	0.03
Two or more psychological risk factors[b]	0.03	0.05
Staff Characteristics		
Race/Ethnicity (vs. white)		
African American	0.03	0.09
Hispanic	-0.10	0.06
Other	0.23**	0.08
Speaks language other than English	0.07	0.06
Has a BA degree or higher	-0.05	0.04
Years of experience in Early Head Start	0.00	0.01
Has a degree in early childhood	0.01	0.05
Has a CDA credential	-0.02	0.05
Program Characteristics		
Multiple Approach[c] (versus single approach)	-0.07	0.05
Population Served		
Over 50% of families with mental health or substance abuse problems	-0.02	0.05
Over 50% of families in unsafe neighborhoods or experiencing family violence	0.00	0.05
Over 50% of families with more than 3 demographic risks	0.05	0.05
Fully Implemented[d]	0.02	0.05
Sample Size (children)	**880**	

Source: Survey Management System (SMS), Family Service Tracking (FST); Staff Child Report; Parent Interview; Program Director Interview.

Note: Estimated using a three-level linear probability model with random effects for program and family. Coefficients indicate differences in the probability of being highly involved associated with each characteristic. For example, holding all other characteristics constant, the probability of being highly involved at age 2 is 8 percentage points lower than the probability at age 1.

[a]The maternal demographic risk is constructed by summing the number of the following risk factors that the mother faced: (1) being a teenage mother, (2) having no high school credential, (3) receiving public assistance, (4) not being employed or in school or training, and (5) being a single mother.
[b]Family psychological risk is based on moderate or severe depressive symptoms, high parenting stress, and current or past substance use problems.
[c]Programs offering both center- and home-based services (as opposed to only one of the 2) are designated as multiple-approach.
[e]Programs with ratings equal to or above 3 on each of the implementation cornerstones are designated as "fully implemented".

+$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$.

CDA = Child Development Associate; DLL=Dual language learner.

Predictors of Service Take-Up Rates

In order to quantify rates of service take-up, we use the percentage of services offered that were used or completed because these variables more clearly capture family participation by adjusting for the role that programs play in offering services. That is, families cannot complete a home visit or attend a center day if programs do not make them available.

The analytic sample for home visit take-up includes children and families who were in the home-based option during one or more years from ages 1 to 3 and did not change service options during the FST data collection period. Likewise, the analytic sample for center take-up includes children who were in the center-based option during one or more years from ages 1 to 3 and did not change service options during the FST data collection period. We include early leavers in the analyses in order to compare their take-up rates with children who stay enrolled through age 3. The take-up rate for early leavers is based on their receipt of services during the weeks that they were enrolled.

The models include the same set of child, family, staff and program characteristics that we used to predict family engagement. For child health status, we use information provided by the parent in the spring immediately preceding the beginning of the FST data collection window in each year. Similarly, for staff characteristics, we use information pertaining to staff members affiliated with each family in the spring preceding the FST window.

Home-Visit Take-Up Rates Are Higher at Younger Ages; Center Take-Up Rates Are Stable Over Time

Home visit take-up rates during the year from ages 1 to 2 were 2 percentage points higher, on average, compared to take-up rates during the year between ages 2 to 3 (Table VIII.3). This means that there is a slight decline in the percentage of home visits that were completed (of those offered) as children got older. We did not find significant differences in center take-up rates by age.

Child and Family Characteristics are Related to Home Visit and Center Take-Up

The child and family characteristics associated with home visit take-up generally differed from those associated with center take-up. The only characteristic that is associated with both home visit and center take-up rates is early leaver status (Table VIII.3). Specifically, families in the home-based option who eventually left their program early have lower home visit take-up rates (12 percentage points lower, on average) compared to families who stayed in the program through age 3. Similarly, families in the center-based option who were early leavers have lower center take-up rates (15 percentage points lower) compared to families who were not early leavers.

In addition to early leaver status, other child and family characteristics are also related to take-up rates, depending on the service option. Compared to white children's families, Hispanic children's families have lower home visit take-up rates (Table VIII.3). African American children's families also have marginally lower home visit take-up rates but the difference only trended toward significance. Families with DLLs have higher home visit take-up rates compared to families who only speak English at home. Center take-up rates were only related to one other child characteristic, gender. Male children have marginally higher rates of take-up compared to female children at a trend level of significance.

Table VIII.3. Predictors of Service Take-Up Rates, by Service Option

	Home Visits[a]		Center Days[a]	
	Coefficient	Standard Error	Coefficient	Standard Error
2 to 3 year[b]	0.02*	0.01	0.02	0.02
Child and Family Characteristics				
Family enrolled in pregnancy	0.04+	0.02	-0.01	0.02
Family left Early Head Start early (before age 3)	-0.12***	0.02	-0.15***	0.04
Race/Ethnicity (vs. white)				
African American	-0.10+	0.05	0.01	0.03
Hispanic	-0.07*	0.03	0.01	0.03
Other	-0.05	0.03	0.04	0.03
Male child	0.02	0.02	0.03+	0.01
DLL	0.08**	0.03	-0.02	0.02
Child born with low or very low birth weight	0.04	0.03	-0.01	0.04
Child in excellent or very good health	0.01	0.02	-0.02	0.02
Maternal risk[c] (vs. lower risk)				
Medium maternal demographic risks	-0.02	0.02	0.00	0.02
High maternal demographic risks	-0.01	0.02	-0.00	0.02
Psychological risk[d] (vs. zero risks)				
One psychological risk factor	-0.02	0.02	-0.02	0.02
Two or more psychological risk factors	0.03	0.03	-0.01	0.03
Staff Characteristics				
Race/Ethnicity (vs. white)				
African American	0.16*	0.07	-0.00	0.02
Hispanic	-0.04	0.05	0.05+	0.03
Other	0.08	0.05	-0.02	0.04
Speaks language other than English	0.04	0.04	-0.02	0.02
Has a BA degree or higher	0.01	0.04	0.05*	0.02
Years of experience in Early Head Start	-0.01	0.01	-0.00	0.00
Has a degree in early childhood	0.03+	0.02	-0.03	0.03
Has a CDA credential	0.03	0.05	-0.01	0.02
Program Characteristics				
Multiple approach[e](versus single approach)	-0.06	0.05	0.02	0.04
Population served				
Over 50% of families with mental health or substance abuse problems	0.01	0.05	0.05+	0.03
Over 50% of families in unsafe neighborhoods or experiencing family violence	-0.01	0.03	-0.03	0.03
Over 50% of families with more than 3 demographic risks	0.02	0.04	0.03	0.03
Fully implemented[f]	0.01	0.04	0.04	0.03
Sample Size (children)	**357**		**370**	

Source: Survey Management System (SMS), Family Service Tracking (FST); Staff Child Report; Parent Interview; Program Director Interview.

Note: Estimated using a three-level model with random effects for program and family. Service take-up rates are calculated by dividing the total number of home visits or center days received by the total number of home visits or center days offered (including services missed due to family reasons, but excluding those missed due to program reasons). Coefficients indicate differences in service take-up rates (in percentage points) associated with each characteristic. For example, the center day take-up rate for families who left the program early was 10 percentage points lower, on average, compared to families who did not leave early.

[a] Sample limited to families who did not change service options.
[b] Coefficient indicates differences in take-up compared to the year in which children were age 1 to 2.
[c] The maternal demographic risk index is constructed by summing the number of the following risk factors that the mother faced: (1) being a teenage mother, (2) having no high school credential, (3) receiving public assistance, (4) not being employed or in school or training, and (5) being a single mother.
[d] Family psychological risk is based on moderate or severe depressive symptoms, high parenting stress, and current or past substance use problems.
[e] Programs offering both center- and home-based services (as opposed to only one of the 2) are designated as multiple-approach.
[f] Programs with ratings equal to or above 3 on each of the implementation cornerstones are designated as "fully implemented".

+$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$.

CDA = Child Development Associate; DLL=Dual language learner.

Staff Race/Ethnicity and Education are Related to Service Take-Up

Compared to families whose home visitors are white, families whose home visitors are African American have higher home visit take-up rates[60] (Table VIII.3). Center take-up rates are also associated with staff education. Children whose teachers have at least a bachelor's degree have higher center take-up rates than children whose teachers have lower levels of education.

Few Program Characteristics are Related to Service Take-Up

We did not find significant associations between program approach or implementation and families' service take-up (Table VIII.3). We found limited evidence of a relationship between center take-up rates and characteristics of the program population. Families in programs serving a high proportion of families with mental health or substance abuse problems have higher center take-up rates compared to families in programs serving fewer families facing such risks, but the difference is only trended toward significance. Home visit take-up rates are not associated with program population characteristics.

Summary of Key Findings

- Participation does not vary substantially between programs.
  - Average length of enrollment is about the same regardless of the particular program that a family attends.
  - Only 7 percent of the variability in engagement is due to program-level differences.
  - Less than a quarter of the variability in home visit completion and center attendance is due to differences between programs.
- Participation varies substantially even among families enrolled in the same program.
  - Seventy-eight percent of the variability in length of enrollment is due to differences between families.
  - Families attending the same program differ considerably in their levels of involvement, with 52 percent of the variability due to differences between families.
  - Families also differ considerably in terms of their service take-up rates.
- Family involvement and service take-up rates vary over time.
  - A family rated as highly involved in one year may not necessarily be rated as such in another year. Forty-one percent of the variation in family involvement is due to differences observed within families over time.
  - Service take-up rates also vary over time, but to a lesser extent. Ten percent of the variability in home visit take-up rates and 16 percent of the variation in center take-up rates is due to year-to-year differences within families.
- Length of enrollment varies by child and family characteristics.

[60] In other model iterations, we tested whether the relationship between staff race and take-up varies by child race (that is, whether the match between race of the home visitor and the child make a difference) but did not find any evidence that this is the case.

  - Children who are DLLs stay in Early Head Start longer compared to children from homes in which English is the only language spoken.
  - There is a trend for children whose mothers face medium demographic risk to be enrolled for a marginally shorter period compared to children facing lower maternal demographic risk.

- Length of enrollment is not related to program approach, population served, or program implementation.
- Family involvement attenuates over time.
  - The probability of being highly involved is lower at ages 2 and 3, compared to age 1.
- Family involvement is related to child and family characteristics.
  - Families who were early leavers have a lower probability of being highly involved compared to families who stay through age 3.
  - Families of Hispanic children are marginally more likely to be highly involved compared to families of white children (at a trend level).
  - Compared to families with low maternal demographic risks, families facing medium or high risks have a lower probability of being highly involved.
- Few staff characteristics are related to family involvement.
- Family involvement is not related to program approach, population served, or program implementation.
- Home visit take-up rates vary by age but center take-up rates do not.
  - Home visit take-up rates during the age 1 to 2 year were higher compared to take-up rates during the age 2 to 3 year.
  - We did not find significant differences in center take-up rates by year.
- Child and family characteristics are related to home visit and center take-up.
  - Compared to families who stay in the program through age 3, families who were early leavers have lower home visit and center take-up rates while they were enrolled.
  - Families who enrolled in pregnancy have marginally higher home visit take-up rates than families who enrolled later (at a trend level).
  - Compared to families of white children, those with Hispanic children have lower home visit take-up rates. African Americans have marginally lower take-up rates (at trend level).
  - Families with DLL children have higher home visit take-up rates compared to families who only speak English at home.
  - Families with male children have marginally higher center take-up rates compared to families with female children (at a trend level).
- Staff race/ethnicity and education are related to service take-up.

  - Compared to families whose home visitors are white, families whose home visitors are African American have higher home visit take-up rates.
  - Families whose home visitor has a degree in early childhood also completed marginally more of the home visits offered to them (at a trend level).
  - Children whose teachers have at least a bachelor's degree have higher center take-up rates than children whose teachers have lower levels of education.

- Few program characteristics are related to service take-up.
  - We did not find significant associations between program approach or implementation and families' service take-up.
  - Families in programs serving a high proportion of families with mental health or substance abuse problems have marginally higher center take-up rates compared to families in programs serving fewer families facing such risks (at a trend level).

IX. RELATIONSHIPS BETWEEN EARLY HEAD START EXPERIENCES AND CHILD, PARENTING, AND FAMILY OUTCOMES

This chapter builds upon the findings presented in earlier chapters that examine Early Head Start participation (Chapter VII) and predictors of participation (Chapter VIII) to build a more comprehensive understanding of the relationships among Early Head Start services and outcomes. The models we construct take advantage of the richness of Baby FACES data and provide useful information to programs, OHS, and ACF for program improvement and planning. The models incorporate child, parenting, and family outcomes at age 3 and relate these to Early Head Start experiences including family participation, service quality, and program implementation, taking into account family and child characteristics. Box IX.1 details the experience and outcome measures we consider; these are also summarized in Chapter II).

Box IX.1. Early Head Start Experiences and Child, Parenting, and Family Outcome Variables

Early Head Start Experiences. The family participation variables we examine are family involvement (as rated by teachers and home visitors), enrollment during pregnancy, length of time in program, and center days attended or home visits received.

Home visit quality is measured by the HOVRS-A Visitor Strategies subscale. We use this subscale as a quality measure, rather than the total score or the Visitor Effectiveness subscale, because it is focused on the visitor herself rather than the visitor-family dyad. Center quality is measured by two subscales of the CLASS-T: Emotional and Behavioral Support and Engaged Support for Learning. Quality and family participation measures are described in more detail in Chapters V and VI.

Program implementation is measured by a dichotomous variable indicating that a child's program was or was not fully implemented in all years. Program implementation ratings are discussed in more detail in Chapter V.

Child, Parenting, and Family Outcomes. We focus on the following child outcomes at age 3 in the language and social-emotional domains: the direct assessments PPVT and PLS-4; the Emotional Regulation subscale of the BRS, rated by observers; and the BITSEA Problem and Competence domains rated by staff (see Chapter III). The parenting outcomes we consider are the Parental Synchronicity composite from the Two-Bags observation, the Parent Support for Child Learning index, which captures the degree of cognitive stimulation provided to children in the home setting, and Parent Provision of Learning Materials index, which is a composite of interview items and observations from the HOME scale (these measures are described in Chapter II). We consider one family economic condition variable as an outcome: level of maternal risk at age 3.

In this chapter, we consider each Early Head Start experience measure separately. We examine bivariate relationships between outcomes of interest and each experience measure and use those raw associations to inform our selection of Early Head Start experience variables to submit to further scrutiny. Box IX.2 contains a description of how we constructed the variables in our analyses, Box IX.3 contains a summary of our modeling approach, and Box IX.4 contains the control variables we use in our multivariate specifications.

Box IX.2. Variable Construction

Some of the Early Head Start experience variables are continuous, such as the dosage and quality measures. Other experience variables are naturally dichotomous, such as whether a mother enrolled during pregnancy and whether a family is rated as consistently highly involved in the program. We created dichotomous versions of the continuous variables to understand if relationships to the outcomes differ at higher versus lower points. We worked with ACF to select thresholds that had policy or programmatic interest and that were possible given the distributions of the variables. That is, if only a few observations had, for example, service uptake of 75 percent or higher (as is true in this case), then that cut point would not be useful as a threshold because there are not enough observations in each category to do the analysis.

Length of Enrollment Thresholds. We examined several different possible cut points that were supported by the distribution of the data. We tested for differences in age 3 outcomes for children on either side of several length-of-enrollment thresholds: 24, 28, 30, and 32 months. The different levels were selected because they represented relatively common lengths of enrollment within the Baby FACES sample. Sixty-five percent of children were enrolled for 24 months or more, 57 percent for 28 months or more, 46 percent for 30 months or more, and 29 percent for 32 months or more.

Center Attendance and Home Visits Received. We measure the total number of center days attended or home visits received over two years (that is, between age 1 and age 3). For the threshold analysis, we defined the thresholds for center attendance and home visits received to be receipt of 50 percent or more of recommended services over two years. For home visits this is 48 out of 96 possible visits. Fifty-three percent of children had 48 or more home visits. For center days the threshold is attending at least 240 center days out of 480. Fifty-seven percent of children attended at least 240 center days.

Home Visit and Classroom Quality. We measure home visit quality using the average of the HOVRS-A Visitor Strategies subscale at ages 2 and 3.[61] For each subscale of the CLASS-T (Emotional and Behavioral Support and Engaged Support for Learning), we take the average across ages 2 and 3. For both home visit and classroom quality, we use the average as a summary measure of a child's experience of quality in Early Head Start. The home visit quality threshold is an average of 3 or higher on the HOVRS-A Visitor Strategies subscale. Sixty-four percent of children experienced average home visit quality at or above this threshold. The classroom quality threshold varies by subscale. The threshold for the Emotional and Behavioral Support subscale is an average score of 5 or higher, with 69 percent of children experiencing average quality on this subscale that exceeds the threshold. The threshold for the Engaged Support for Learning subscale is an average score of 3 or higher over two years (average scores are lower and therefore require a lower threshold); 61 percent of children experience quality on this subscale that exceeds the threshold.

[61] Visitor Strategies scores are averaged across two years (ages 2 and 3, using multiple imputation to fill in missing scores).

Box IX.3. Modeling Approach

Baseline, Experience, and Outcome Variables. We consider relationships between age 3 outcomes and Early Head Start experience variables measured at ages 2 and 3 (spring 2010 and spring 2011 for the 1-year-old Cohort, and spring 2011 and spring 2012 for the Newborn Cohort). Characteristics measured at age 1 (spring 2009 for the 1-year-old Cohort and spring 2010 for the Newborn Cohort) are used as "baseline" controls.

Bivariate Analyses as a First Step. As a first step in analysis, we assess the bivariate relationship (that is, without controlling for other variables) between each experience and outcome variable using pairwise correlations for the continuous variables (e.g., attendance and quality measures), and *t*-tests of group mean differences for the dichotomous experience variables (e.g., enrollment during pregnancy and staff rated involvement). In this way we can assess whether there appears to be any relationship between the experience and outcome variables. In addition, we attempt to determine if there are threshold effects for the continuous variables. That is, we examine different thresholds using *t*-tests of group mean differences to determine if participation above a particular threshold has a different relationship with outcomes than participation below that threshold. If we find significant bivariate relationships, we proceed with multivariate analyses since bivariate analyses do not include control variables that allow us to determine whether the apparent relationship is driven by another variable. Multivariate analysis allows a more nuanced understanding of the relations between experiences and outcomes in the context of other variables.

Multivariate Analyses When Warranted. We analyze each continuous outcome that has a significant bivariate association (*p*-value of 0.05 or less) with an Early Head Start experience measure as the dependent variable in a multivariate linear regression. In each regression, we include a large set of controls that include the child, family, staff, and program characteristics described in Box IX.4. The coefficient of interest in these regressions is that associated with the Early Head Start experience variable which describes the relationship between the experience measure and the outcome. For the dichotomous experience variables (including those we create as part of threshold analyses), we conceptualize the Early Head Start experience variables as "treatments," where the treatment is experiencing high levels of participation, quality, or program implementation. In order to measure the relationships between these treatments and outcomes, we use multivariate regression, and as a sensitivity analysis use a technique called doubly robust estimation. This technique combines propensity score methods with regression in order to remove bias and increase efficiency (Schafer and Kang 2008).

The Doubly Robust Estimation Technique. Doubly robust estimation involves two steps. First, we create propensity scores by estimating an individual's likelihood of experiencing a high level of participation, quality, or program implementation based on age 1 child, family, and program characteristics. We use the inverse of these propensity scores as weights in a weighted multivariate linear regression. "Doubly robust" means that, under certain strong assumptions and as long as either the propensity model or the regression model is specified correctly, the estimate has a causal interpretation (Schafer and Kang 2008). We use this method as a check that our multivariate linear regression results are not sensitive to the models' assumptions about the nature of the relationship between control variables and outcomes (i.e., functional form assumptions).

The Problem of Omitted Variables. The most important assumption that must hold for multiple linear as well as doubly robust regression to estimate a causal effect is the "no unobservable confounders" (or no omitted variables) assumption. Having no unobservable confounders means that all variables that predict participation, quality, or implementation and that are relevant for the outcome are included in the model. This is both a strong and an untestable assumption. Baby FACES offers a uniquely rich set of data sources, but even with this wealth of data we cannot say with certainty that we have controlled for all relevant variables or that the results we report warrant a causal interpretation.

Box IX.4. Control Variables

- ***Basic child characteristics***
  - Cohort
  - Race/ethnicity
  - Age as of spring 2009
  - Gender
  - DLL status
- ***Child health***
  - Low birth weight
  - Child's health (parent-reported), age 1
- ***Baseline child outcomes***
  - Parent-reported ASQ Communication, Problem Solving, and Personal Social
- ***Baseline parenting and family characteristics***
  - Parent depressive symptoms
  - Parenting Stress Index
  - High family involvement (staff report)
  - Maternal risk composite (total number of five risk factors: family receiving public assistance; mother not in school, training, or employment; mother has less than a high school education; single-parent family; teen mother status)
- ***Service type***
  - Home visiting only
  - Center only
  - Other (includes children in the combination option as well as those experiencing service changes)
- ***Program characteristics***
  - Multiple approach
  - Urbanicity
  - Total enrollment
  - Percentage of families speaking Spanish
  - Region
  - Majority of families served by program have mental health or substance abuse problems
  - Majority of families served by program reside in unsafe neighborhoods or experience family violence
  - Majority of families served by program have more than three demographic risks

Children in Highly Involved Families Have Better Behavioral Outcomes at Age 3

Parental involvement is an important measure of family participation. Previous research based on the EHSREP has shown that staff-rated overall engagement is associated with outcomes such as maternal depression (Raikes et al. 2006). Here we consider the relationships between consistently high family involvement and outcomes. A family is considered consistently highly involved if the family received a staff rating of four, the highest possible score, on a family involvement scale over two years (when the child was 2 and 3).

Table IX.1 contains the results of the bivariate analyses and compares the mean of each outcome across two different types of families: those rated as consistently highly involved and those not rated as such. At age 3, children from highly involved families have significantly higher PLS-4 and BITSEA

Competence Domain scores, and significantly lower BITSEA Problem Domain scores. In addition, highly involved families have fewer maternal risks. However, the two types of families do not differ in terms of PPVT-4 scores or parenting outcomes.

Table IX.1. Means of Age 3 Child, Parenting, and Family Outcomes by Level of Involvement

	Not Consistently Highly Involved		Consistently Highly Involved		
	Mean	Standard Error	Mean	Standard Error	**Sample Size**
PPVT-4	90.8	1.15	91.8	1.99	**455**
PLS-4 (English)*	95.0	1.50	100.4	2.10	**402**
PLS-4 (Spanish)	95.5	1.71	100.0	2.78	**193**
PLS-4 (Bilingual)	104.2	2.33	107.0	3.34	**199**
BRS Emotional Regulation**	40.4	0.63	42.7	0.79	**543**
BITSEA Problem Domain (Staff Report)***	7.8	0.40	5.3	0.56	**544**
BITSEA Competence Domain (Staff Report)***	16.2	0.19	17.7	0.30	**542**
Two-Bags Parental Synchronicity	4.4	0.06	4.5	0.10	**546**
Parent Support for Child Learning Index	0.0	0.04	0.1	0.05	**544**
Parent Provision of Learning Materials Index	0.0	0.05	0.0	0.07	**548**
Age 3 Maternal Risk**	2.5	0.08	2.0	0.16	**553**

Sources: Parent Interview, Staff-Child Report, Direct Child Assessment, Home Visit Observation, Survey Management System.

Note: Results calculated using multiple imputed data. Results from bivariate analyses using no controls reported.

+$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$.

PPVT = Peabody Picture Vocabulary Test, PLS = Preschool Language Scale, BRS = Behavioral Rating Scale, BITSEA = Brief Infant Toddler Social Emotional Assessment

Since we observed statistically significant bivariate relationships, we proceeded with a multivariate analysis of the relationship between consistently high involvement and outcomes. Table IX.2 contains results from two regression models: a linear model and a doubly robust model incorporating inverse probability weights and regression (see Box IX.2 for covariates used and Appendix E for full regression results). Results from both models are similar. Controlling for a large set of child, family, and program characteristics—including staff-rated family involvement at age 1—we find that children in highly involved families exhibit significantly fewer negative and significantly more positive behaviors as rated by their teacher or home visitor. Children in these families have better emotional regulation capabilities as rated by observers, although this relationship was only statistically significant in the regular linear regression analysis and trended toward significance in the doubly robust model. Though not statistically significant, there is a trend in the doubly robust regression model to suggest that highly involved families also tend to have fewer maternal risks.

Table IX.2. Comparing Two Different Methods to Measure Relationships Between Level of Involvement and Age 3 Outcomes

	Linear Regression	Doubly Robust	
	Coefficient on High Involvement	Coefficient on High Involvement	Sample Size
PLS-4 (English)	2.67	1.54	360
BRS Emotional Regulation	1.77*	1.47+	489
BITSEA Problem Domain (Staff Report)	-1.77***	-1.97***	485
BITSEA Competence domain (Staff Report)	1.39***	1.71***	484
Age 3 Maternal Risk	-0.25**	-0.16+	517

Sources: Parent Interview, Staff-Child Report, Direct Child Assessment, Home Visit Observation, Survey Management System.

Note: Results calculated using multiple imputed data and with child, family, and program characteristics as controls. See Box IX.4 for a list of control variables.

+p < .10; *p < .05; **p < .01; ***p < .001.

PLS = Preschool Language Scale, BRS = Behavioral Rating Scale, BITSEA = Brief Infant Toddler Social Emotional Assessment.

Enrollment During Pregnancy Is Not Related to Outcomes After Controlling for Various Characteristics

The EHSREP found larger impacts for families who enrolled during pregnancy on a number of parenting and child cognitive and social-emotional development outcomes, compared to families who enrolled after the study's focus child was born (ACF 2002a).

Table IX.3 contains bivariate analyses of the relationship between enrollment during pregnancy and our outcomes of interest. The table shows that 3-year-olds whose families enrolled during pregnancy have significantly lower mean PPVT-4 scores and significantly fewer maternal risks, on average. When controlling for the child, family, and program characteristics listed in Box IX.2, however, these relationships cease to be statistically significant. (See Appendix E for multivariate regression results.) Enrollment during pregnancy was unrelated to PLS scores, BITSEA, and parenting outcomes in the bivariate analyses.

Length of Enrollment Is Not Associated with Age 3 Outcomes After Controlling for Various Characteristics

Variation in length of enrollment can come from two sources: differences in when families enrolled in Early Head Start, and differences in when families left Early Head Start. Since children who have data on age 3 outcomes are not early exiters, the only source of variation in length of enrollment comes from differences in when families initially enrolled. For children with age 3 outcomes, length of enrollment has relatively low variability, making it difficult to measure relationships between outcomes and length of enrollment. As Table IX.4 shows, we find no significant relationships between length of enrollment and child, parenting, and family outcomes.

Table IX.3. Means of Age 3 Child, Parenting, and Family Outcomes by Pregnancy Enrollment Status

	Enrolled After Pregnancy		Enrolled During Pregnancy		
	Mean	Standard Error	Mean	Standard Error	Sample Size
PPVT-4*	92.6	1.05	88.8	1.68	455
PLS-4 (English)	95.9	1.37	98.4	2.25	402
PLS-4 (Spanish)	95.9	2.08	99.4	2.82	193
PLS-4 (Bilingual)	104.1	2.28	107.0	3.17	199
BRS Emotional Regulation	41.4	0.64	41.0	0.96	543
BITSEA Problem Domain (Staff Report)	7.3	0.49	6.4	0.66	544
BITSEA Competence Domain (Staff Report)	16.6	0.18	16.9	0.31	542
Two-Bags Parental Synchronicity	4.4	0.07	4.5	0.09	546
Parent Support for Child Learning Index	0.0	0.03	0.0	0.07	544
Parent Provision of Learning Materials Index	0.0	0.05	0.0	0.07	548
Maternal Risk*	2.4	0.10	2.1	0.15	553

Sources: Parent Interview, Staff-Child Report, Direct Child Assessment, Home Visit Observation, Survey Management System.

Note: Results calculated using multiple imputed data. Results from bivariate analyses using no controls reported.

+$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$.

PPVT = Peabody Picture Vocabulary Test, PLS = Preschool Language Scale, BRS = Behavioral Rating Scale, BITSEA = Brief Infant Toddler Social Emotional Assessment

Table IX.4. Pairwise Correlations Between Age 3 Outcomes and Length of Enrollment

	Length of Enrollment (Months)	Sample Size
PPVT-4	-0.24+	455
PLS-4 (English)	0.06	402
PLS-4 (Spanish)	0.16	193
PLS-4 (Bilingual)	0.21	199
BRS Emotional Regulation	0.04	543
BITSEA Problem Domain (Staff Report)	-0.15	544
BITSEA Competence Domain (Staff Report)	0.07	542
Two-Bags Parental Synchronicity	-0.02	546
Parent Support for Child Learning Index	0.01	544
Parent Provision of Learning Materials Index	-0.04	548
Maternal Risk	-0.18	553

Sources: Parent Interview, Staff-Child Report, Direct Child Assessment, Home Visit Observation, Survey Management System.

Note: Results calculated using multiply imputed data.

+$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$.

PPVT = Peabody Picture Vocabulary Test, PLS = Preschool Language Scale, BRS = Behavioral Rating Scale, BITSEA = Brief Infant Toddler Social Emotional Assessment.

We tested for differences in age 3 outcomes for children on either side of several common length-of-enrollment thresholds: 24, 28, 30, and 32 months (see Box IX.2). We observe one significant difference in these threshold analyses: children who were enrolled for longer than 30 months score significantly higher on the bilingual PLS-4 than children enrolled for 30 months or less (mean difference = 6.0, *p*-value = 0.02). When the full set of controls is included, however, this difference ceases to be significantly different from zero. (See Appendix E for multivariate regression results.)

Average Home Visit Quality Is Not Related to Age 3 Outcomes, but Quality Thresholds May Matter for Spanish Language Development

Measures to assess home visit quality such as the HOVRS-A have only recently been developed, and research on associations between home visit quality and outcomes is scant. Table IX.5 shows pairwise correlations between the age 3 outcomes of interest and average quality measured at ages 2 and 3. Average home visit quality does not appear to be related to any of the age 3 outcomes of interest.

Table IX.5. Pairwise Correlations Between Age 3 Outcomes and Average Home Visit Quality (HOVRS-A Visitor Strategies), Ages 2-3

	Average Home Visit Quality	Sample Size
PPVT-4	-0.02	187
PLS-4 (English)	-0.03	180
PLS-4 (Spanish)	0.23	118
PLS-4 (Bilingual)	0.20	119
BRS Emotional Regulation	-0.04	243
BITSEA Problem Domain (Staff Report)	-0.07	223
BITSEA Competence Domain (Staff Report)	-0.06	225
Two-Bags Parental Synchronicity	0.06	272
Parent Support for Child Learning Index	0.04	327
Parent Provision of Learning Materials Index	-0.04	272
Maternal Risk	0.12	307

Sources: Parent Interview, Staff-Child Report, Direct Child Assessment, Home Visit Observation, Survey Management System.

Note: Results calculated using multiple imputed data. Sample limited to children in the home visit option who did not change service type. There are no significant correlations.

PPVT = Peabody Picture Vocabulary Test, PLS = Preschool Language Scale, BRS = Behavioral Rating Scale, BITSEA = Brief Infant Toddler Social Emotional Assessment.

We conducted threshold analyses to determine if outcomes differ for children who experienced home visits that were, on average, of adequate quality or better (i.e., an average score of 3 or more on the HOVRS-A Visitor Strategies 5-point subscale), compared to children experiencing lower-quality home visits. We observe one significant difference in our bivariate analysis: children with higher-quality home visits had significantly higher Spanish PLS-4 scores (mean difference = 12.7, *p*-value = 0.04). This difference remains when the full set of controls is added (regression coefficient = 14.18, *p*-value = 0.04).[62] (See Appendix E for multivariate regression results.)

[62] Due to sample size limitations, doubly robust estimation with the full set of controls is infeasible for this analysis. However, doubly robust estimation with a restricted set of covariates yields similar results. See Appendix E.

Modest Evidence Suggests that Center Quality Has Positive Associations with Age 3 Language Outcomes

Previous research has shown that attendance in a high quality child care center has a positive association with language development (Mashburn et al. 2008). One reason for this association may be that classroom quality itself affects language learning. Table IX.6 shows that the CLASS-T Emotional and Behavioral Support subscale is negatively correlated with BITSEA Problem domain scores, and positively correlated with PPVT-4 and English PLS-4 scores. However, when the full set of child, family, and program characteristics are used as controls, the relationships become insignificant. (See Appendix E.)

We test for differences across an average quality threshold of 5 on the CLASS-T Emotional and Behavioral Support 7-point subscale. Children above this threshold experienced medium-to-high quality center care at ages 2-3. We find no statistically significant differences in any of the outcomes with respect to this threshold.[63]

Table IX.6. Pairwise Correlations Between Age 3 Outcomes and Average Center Quality (CLASS-T Emotional and Behavioral Support), Ages 2-3

	Average Center Quality	Sample Size
PPVT-4	0.19*	269
PLS-4 (English)	0.21*	239
PLS-4 (Spanish)	-0.04	99
PLS-4 (Bilingual)	0.10	102
BRS Emotional Regulation	0.10	286
BITSEA Problem Domain (Staff Report)	-0.20*	261
BITSEA Competence Domain (Staff Report)	0.11	257
Two-Bags Parental Synchronicity	0.07	305
Parent Support for Child Learning Index	0.19+	343
Parent Provision of Learning Materials Index	-0.01	296
Maternal Risk	-0.15	321

Sources: Parent Interview, Staff-Child Report, Direct Child Assessment, Home Visit Observation, Survey Management System.

Note: Results calculated using multiple imputed data. Sample limited to children in the center-based option who did not change service type.

+$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$.

PPVT = Peabody Picture Vocabulary Test, PLS = Preschool Language Scale, BRS = Behavioral Rating Scale, BITSEA = Brief Infant Toddler Social Emotional Assessment.

[63] We find significant differences across a threshold of 6 in language development. However few children experience Emotional and Behavioral Support of 6 or greater, the data do not support multivariate analyses with this threshold. See Appendix E.

We also consider the CLASS-T Engaged Support for Learning subscale in relation to age 3 outcomes. Table IX.7 shows that it is positively correlated with language development. The relationship to the PPVT-4 remains statistically significant when the full set of controls is used (Table IX.8).

Table IX.7. Pairwise Correlations Between Age 3 Outcomes and Average Center Quality (CLASS-T Engaged Support for Learning), Ages 2-3

	Average Center Quality	Sample Size
PPVT-4	0.28***	269
PLS-4 (English)	0.23*	239
PLS-4 (Spanish)	0.12	99
PLS-4 (Bilingual)	0.37***	102
BRS Emotional Regulation	0.12+	286
BITSEA Problem Domain (Staff Report)	-0.05	261
BITSEA Competence Domain (Staff Report)	0.06	257
Two-Bags Parental Synchronicity	0.08	305
Parent Support for Child Learning Index	0.11	343
Parent Provision of Learning Materials Index	0.07	296
Maternal Risk	-0.19	321

Sources: Parent Interview, Staff-Child Report, Direct Child Assessment, Home Visit Observation, Survey Management System.

Note: Results calculated using multiple imputed data. Sample limited to children in the center-based option who did not change service type.

+p < .10; *p < .05; **p < .01; ***p < .001.

PPVT = Peabody Picture Vocabulary Test, PLS = Preschool Language Scale, BRS = Behavioral Rating Scale, BITSEA = Brief Infant Toddler Social Emotional Assessment.

Table IX.8. Relationships Between Age 3 Outcomes and Average Center Quality (Engaged Support for Learning), Ages 2-3

	Average Center Quality	Sample Size
PPVT-4	2.76*	244
PLS-4 (English)	2.49+	215
PLS-4 (Bilingual)	4.0	99

Sources: Parent Interview, Staff-Child Report, Direct Child Assessment, Home Visit Observation, Survey Management System.

Note: Results calculated using multiple imputed data and with child, family, and program characteristics as controls. See Box IX.4 for a list of control variables. Sample limited to children in the center-based option who did not change service type.

+p < .10; *p < .05; **p < .01; ***p < .001.

PPVT = Peabody Picture Vocabulary Test, PLS = Preschool Language Scale.

For the CLASS-T Engaged Support for Learning subscale, we test for differences across an average quality threshold of 3, indicating medium-quality center care.[64] Table IX.9 shows that there are significant differences on the PPVT-4 and the bilingual PLS-4. When the full set of controls are included in multivariate linear and doubly robust regressions, however, these differences are not significantly different from zero. (See Appendix E.)

[64] As discussed in Chapter V, ratings for Engaged Support for Learning tend to be lower than for Emotional and Behavioral Support.

Table IX.9. Means of Age 3 Child, Parenting, and Family Outcomes Across a Medium Center Quality Threshold (Average Score of 3 or Higher on the CLASS-Engaged Support for Learning Subscale at Ages 2-3)

	Below Threshold		Above Threshold		
	Mean	Standard Error	Mean	Standard Error	Sample Size
PPVT-4*	91.0	1.58	95.8	2.11	269
PLS-4 (English)	98.1	2.52	100.9	3.06	239
PLS-4 (Spanish)	89.0	5.61	95.5	5.78	99
PLS-4 (Bilingual)**	98.5	4.17	112.1	4.67	102
BRS Emotional Regulation	42.5	0.76	43.6	0.97	286
BITSEA Problem Domain (Staff Report)	8.8	1.00	8.1	1.16	261
BITSEA Competence Domain (Staff Report)	16.4	0.45	17.2	0.65	257
Two-Bags Parental Synchronicity	4.3	0.14	4.5	0.16	305
Parent Support for Child Learning Index	0.0	0.08	0.0	0.08	343
Parent Provision of Learning Materials Index	0.1	0.10	0.0	0.11	296
Maternal Risk+	2.4	0.15	2.1	0.20	321

Sources: Parent Interview, Staff-Child Report, Direct Child Assessment, Home Visit Observation, Survey Management System.

Note: Results calculated using multiple imputed data and with child, family, and program characteristics as controls. See Box IX.4 for a list of control variables. Sample limited to children in the center-based option who did not change service type.

+$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$.

PPVT = Peabody Picture Vocabulary Test, PLS = Preschool Language Scale, BRS = Behavioral Rating Scale, BITSEA = Brief Infant Toddler Social Emotional Assessment.

Program Implementation Is Not Related to Age 3 Outcomes

The EHSREP (ACF 2002a; Love et al. 2005) found that impacts were largest for mixed-approach programs that fully implemented Head Start Program Performance Standards early. In the current, non-experimental study, we find little evidence of a relationship between our measure of program implementation, described in Chapter V, and outcomes. Table IX.10 shows that there are no significant differences between the means of outcomes for children in programs that were fully implemented compared to outcomes for children in programs that were not fully implemented.

The Number of Home Visits Received Is Not Related to Age 3 Outcomes After Controlling for Various Characteristics

Previous research on the effect of the number of home visits on child, family, and parenting outcomes is mixed. In a meta-analysis, Sweet and Appelbaum (2004) find that home visit frequency is positively related to child cognitive outcomes. Raikes et al. (2006) find no relationship with child development outcomes, but a positive relationship with the home language and literacy environment.

Table IX.10. Means of Age 3 Child, Parenting, and Family Outcomes by Program Implementation Status

	Not Fully Implemented		Fully Implemented		
	Mean	Standard Error	Mean	Standard Error	Sample Size
PPVT-4	92.6	1.47	90.3	1.84	455
PLS-4 (English)	97.5	2.10	96.4	2.49	402
PLS-4 (Spanish)	94.2	3.94	98.3	4.45	193
PLS-4 (Bilingual)	102.7	2.71	106.2	3.49	199
BRS Emotional Regulation	41.3	0.79	41.2	0.96	543
BITSEA Problem Domain (Staff Report)	6.5	0.48	7.2	0.66	544
BITSEA Competence Domain (Staff Report)	16.7	0.37	16.7	0.41	542
Two-Bags Parental Synchronicity	4.4	0.10	4.4	0.12	546
Parent Support for Child Learning Index	0.0	0.05	0.0	0.07	544
Parent Provision of Learning Materials Index	0.1	0.08	0.0	0.10	548
Maternal Risk	2.4	0.16	2.2	0.18	553

Sources: Parent Interview, Staff-Child Report, Direct Child Assessment, Home Visit Observation, Survey Management System.

Note: Results calculated using multiple imputed data. None of the differences are significant.

PPVT = Peabody Picture Vocabulary Test, PLS = Preschool Language Scale, BRS = Behavioral Rating Scale, BITSEA = Brief Infant Toddler Social Emotional Assessment

Table IX.11 shows pairwise correlations between the child, parenting, and family outcomes we consider and the number of home visits a child received between ages 1 and 3. The total number of home visits received does not appear to have a strong relationship to child or parenting outcomes. Correlations are generally low and not significantly different from zero. One exception is the BRS Emotional Regulation measure. The total number of home visits is negatively correlated with this measure. The relationship becomes non-significant once the full set of child, family, and program controls are added. (See Appendix E.)

We also tested to determine if children who received greater than 50 percent of recommended home visits over two years (i.e., at least 48 out of 96) have better outcomes than children receiving fewer than 50 percent of home visits. We do not observe any significant differences using this threshold.

Limited Evidence to Suggest that Center Attendance Exhibits Positive Associations with Age 3 Language Outcomes

Previous research has shown that child care center attendance is positively related to language development, particularly for low-income children (Caughy et al. 1994). The pairwise correlations in Table IX.12 show that the total number of center days attended is positively correlated with language outcomes (the PPVT-4 and the English PLS-4). Families attending more center days also have higher scores on the Parent Support for Child Learning index. However, when the full set of controls is used, the relationship with these outcomes ceases to be statistically significant. (See Appendix E for results.)

Table IX.11. Pairwise Correlations Between Age 3 Outcomes and Total Home Visits Received, Age 1-3

	Total Home Visits	Sample Size
PPVT-4	-0.01	188
PLS-4 (English)	-0.07	176
PLS-4 (Spanish)	0.07	116
PLS-4 (Bilingual)	0.05	118
BRS Emotional Regulation	-0.25	245
BITSEA Problem Domain (Staff Report)	-0.01	229
BITSEA Competence Domain (Staff Report)	-0.06	231
Two-Bags Parental Synchronicity	0.12	271
Parent Support for Child Learning Index	0.04	324
Parent Provision of Learning Materials Index	-0.09	273
Maternal Risk	-0.01	303

Sources: Parent Interview, Staff-Child Report, Direct Child Assessment, Home Visit Observation, Survey Management System, Family Services Tracking System.

Note: Results calculated using multiple imputed data. Sample limited to children in the home visiting option who did not change service type. None of the differences are significant.

PPVT = Peabody Picture Vocabulary Test, PLS = Preschool Language Scale, BRS = Behavioral Rating Scale, BITSEA = Brief Infant Toddler Social Emotional Assessment

Table IX.12. Pairwise Correlations Between Age 3 Outcomes and Total Center Days Attended, Age 1-3

	Total Center Days	Sample Size
PPVT-4	0.23*	243
PLS-4 (English)	0.45***	214
PLS-4 (Spanish)	-0.13	92
PLS-4 (Bilingual)	0.23	95
BRS Emotional Regulation	0.03	260
BITSEA Problem Domain (Staff Report)	0.00	237
BITSEA Competence Domain (Staff Report)	0.11	234
Two-Bags Parental Synchronicity	0.05	278
Parent Support for Child Learning Index	0.22*	315
Parent Provision of Learning Materials Index	0.20+	270
Maternal Risk	-0.25+	294

Sources: Parent Interview, Staff-Child Report, Direct Child Assessment, Home Visit Observation, Survey Management System, Family Services Tracking System.

Note: Results calculated using multiple imputed data. Sample limited to children in the center-based option who did not change service type.

+$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$.

PPVT = Peabody Picture Vocabulary Test, PLS = Preschool Language Scale, BRS = Behavioral Rating Scale, BITSEA = Brief Infant Toddler Social Emotional Assessment

We tested to determine if children who attended at least 50 percent of recommended center days over two years (i.e., at least 240 out of 480) have better outcomes than those attending fewer than 50 percent of recommended center days. We observe three significant differences: children attending at least 50 percent of recommended center days have higher PPVT-4, English PLS-4, and Parent Support for Child Learning index scores (Table IX.13). In a multivariate linear regression including the full set of controls, the relationship between this threshold and English PLS-4 scores remains significant. Though sizable and positive, the relationship is not significant using the doubly robust method (Table IX.14. (See Appendix E for full regression results.)

Table IX.13. Means of Age 3 Child, Parenting, and Family Outcomes Across a Center Attendance Threshold (Attended at Least 50 Percent of Recommended Days)

	Attended Fewer than 50% of Recommended Days		Attended at Least 50% of Recommended Days		
	Mean	Standard Error	Mean	Standard Error	Sample Size
PPVT-4*	88.1	2.66	95.2	3.07	243
PLS-4 (English)*	90.6	4.07	101.3	4.21	214
PLS-4 (Spanish)	97.6	3.54	94.0	4.42	92
PLS-4 (Bilingual)	104.7	3.70	110.0	5.67	95
BRS Emotional Regulation	43.4	1.06	43.2	1.17	260
BITSEA Problem Domain (Staff Report)	7.8	1.24	8.6	1.46	237
BITSEA Competence Domain (Staff Report)	16.8	0.62	16.9	0.72	234
Two-Bags Parental Synchronicity	4.3	0.19	4.5	0.23	278
Parent Support for Child Learning Index*	-0.1	0.07	0.1	0.08	315
Parent Provision of Learning Materials Index	-0.1	0.11	0.1	0.14	270
Maternal Risk	2.5	0.20	2.1	0.26	294

Sources: Parent Interview, Staff-Child Report, Direct Child Assessment, Home Visit Observation, Survey Management System, Family Services Tracking System.

Note: Results calculated using multiple imputed data. Sample limited to children in the center-based option who did not change service type. Results from bivariate analyses using no controls reported.

+$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$.

PPVT = Peabody Picture Vocabulary Test, PLS = Preschool Language Scale, BRS = Behavioral Rating Scale, BITSEA = Brief Infant Toddler Social Emotional Assessment.

Table IX.14. Comparing Two Different Methods to Measure Relationships Between a Center Attendance Threshold (Attended at Least 50 Percent of Recommended Days) and Age 3 Outcomes

	Linear Regression	Doubly Robust	
	Coefficient on at Least 50% Attendance	Coefficient on at Least 50% Attendance	Sample Size
PPVT-4	5.85+	5.89	219
PLS-4 (English)	9.07*	5.46	190
Parent Support for Child Learning Index	0.18+	0.17	297

Sources: Parent Interview, Staff-Child Report, Direct Child Assessment, Home Visit Observation, Survey Management System, Family Services Tracking System.

Note: Results calculated using multiple imputed data and with child, family, and program characteristics as controls. See Box IX.4 for a list of control variables. Sample limited to children in the center-based option who did not change service type.

+$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$.

PLS = Preschool Language Scale, BRS = Behavioral Rating Scale, BITSEA = Brief Infant Toddler Social Emotional Assessment.

Summary of Key Findings

- Children in highly involved families have better behavioral outcomes at age 3.
  - Controlling for a large set of child, family, and program characteristics children in highly involved families exhibit significantly fewer negative and significantly more positive behaviors as rated by their teacher or home visitor. Children in these families have better emotional regulation capabilities as rated by observers.
- Enrollment during pregnancy is not related to outcomes.
- Length of enrollment is not associated with age 3 outcomes.
- Average home visit quality is not related to age 3 outcomes, but quality thresholds may matter for Spanish language development.
  - Spanish-speaking children with higher-quality home visits (i.e., those with an average score of 3 or more on the HOVRS-A Visitor Strategies subscale) had significantly higher Spanish PLS-4 scores at age 3 after controlling for background characteristics.
- Center quality exhibits positive associations with age 3 language outcomes.
  - We find a positive, statistically significant relationship between the CLASS-T Engaged Support for Learning subscale and PPVT-4 scores at age 3 after controlling for background characteristics.
- Program implementation is not related to age 3 outcomes.
  - There are no significant differences between the means of outcomes for children in programs that were fully implemented compared to outcomes for children in programs that were not fully implemented.
- The number of home visits received is not related to age 3 outcomes.

- There is limited evidence that center attendance exhibits positive associations with age 3 language outcomes.
  - Bivariate analyses of a center attendance threshold (attended at least 50 percent of recommended center days) show positive associations with language and parenting outcomes. The association with English PLS-4 scores remains significant in a multiple linear regression with the full set of controls, but not in a doubly robust regression.

X. BABY FACES OVERARCHING THEMES

This closing chapter draws attention to findings and themes of importance, and reviews the limitations of the data and the research design.

The Kids Are (Mostly) Alright

Baby FACES collected information on children's skills and behavior through multiple modes and reporters including parent and teacher/home visitor reports, direct child assessments, and video-recorded parent-child interactions at ages 2 and 3. These rich data show that children are making good developmental progress in some areas, but are behind norms in others. Although children generally have positive home environments, these are frequently located in disadvantaged neighborhoods.

- Children who remain in Early Head Start until age 3 are approaching norms on some measures of development, although not all. On a measure of auditory comprehension (PLS-4) children are performing similar to norms, but performing less well on receptive vocabulary (measured by the PPVT-4). On average, children's expressive language (on the ECI) is lower than norms, although the measurement approach used may have depressed scores somewhat.[65] Less clear are the findings on the parent and staff reported measure of language development (CDI), which paint a mixed picture when considering different reporters. Parents report better expressive language than teachers or home visitors.
- According to parent report, children are in good overall health and most have insurance coverage and access to a regular source of health care. However, children's height and weight during annual data collection visits show that overweight and obesity are a growing concern. Similar to national averages, about one-third of the children are overweight or obese by age 3.
- Children's parents are providing positive support for their development (based on observations by the assessors in the home, items from the HOME inventory, and ratings of a video-recorded semi-structured parent-child play activity). Unfortunately, the neighborhoods in which children live are characterized by deteriorated housing units, prevalence of trash or garbage, heavy traffic, and a generally unsafe atmosphere.

Staff Bring a Variety of Strengths, but Face Challenges to Providing High Quality Services

Across all years of the study, classrooms and home visit quality are in the mid-range, with most children experiencing a fairly narrow range of service quality and few receiving high or very low quality services. However, staff have the education, experience, positive outlook on their jobs, and positive relationships with parents that would be expected to support quality practice.

- Home visitors and teachers alike display skill in the emotionally supportive aspects of the services they provide. For example, observations of home visits show that visitors generally have positive relationships with the families they serve, and are rated as being most successful in aspects of engaging the family in the visit (HOVRS-A Visitor Effectiveness). Somewhat similarly, classroom observations show that teachers are most

[65] Children in the sample completed this task with the assessor as their play partner; as a relatively unfamiliar adult, the child may have been less verbally expressive than with a more familiar person. Also, by recording and coding the video, any unclear verbalizations could have been "undercounted" as utterances.

successful at providing a warm and positive climate that provides emotional support for children, with few negative interactions (CLASS-T Emotional and Behavioral Support).

- In both cases, home visitors and teachers face challenges related to some aspects of service provision.
  - For home visitors, ratings of the visitor's ability to be responsive to the family, develop a positive relationship, and facilitate parent-child interaction without being intrusive show that these areas are more challenging.
  - For teachers, aspects of instruction are challenging, particularly modeling language use and providing feedback to children (for example, providing hints or even physical assistance as a child works to complete a puzzle [scaffolding] and asking questions or making statements in response to a child's actions with the intention of helping children learn). The relatively low levels of instructional support observed in Baby FACES are similar to findings in other studies about the quality of toddler classrooms, which underscores that implementing these aspects of quality at higher levels is challenging generally.

- Analyses of predictors of service quality do not provide many clear and actionable targets for improving home visit quality (that is, home visit quality is not related to home visitor education or experience or mental health), although job satisfaction is a fairly strong predictor. Classroom quality is somewhat more associated with teacher characteristics (such as education and mental health) and other features such as the quality of the relationship with parents, and job satisfaction.
  - The factors that are important in home visits continue to be a puzzle. Until we have a better sense of the components of a home visit that are most important for effecting change and a way to measure them, the primary staff characteristic that relates to quality is job satisfaction.
  - For classroom quality, the findings suggest a few potential ways that programs can explore to support and enhance quality, such as developing strategies to enhance staff-parent relationships and job satisfaction, supporting teachers who report depressive symptoms, and supporting continuing education. Targeted work exploring the strategies used by programs with high levels of staff morale and job satisfaction could be one avenue for future research.

Families' Program Participation Is Relatively High

Baby FACES collected information on length of enrollment of families, and for the first time on a large scale, information about the services that programs offer and families' participation in those services. We were also able to look at predictors of participation.

- Those who remain in the program throughout eligibility take up a large proportion of the services that are offered (about 75 to 77 percent of home visits and 85-86 percent of center days at ages 2 and 3). This equates to about 3 home visits per month and about 4 center days per week. Families of DLL children tend to have longer enrollment and higher home visit take-up rates than others. Families with lower demographic risk tend to have higher ratings of family involvement. Although service receipt is somewhat lower than recommendations by OHS, the amount of services most families receive is still appreciable. Future work to better understand what amount of services is needed for

change, and perhaps whether different families benefit from different intensity of services could inform policy and practice.

- Families remain in the program an average of just over two years (28 months). However, rates of attrition are relatively high, with 37 percent of families leaving the program before children turn 3. Length of stay among these families is on average just under a year and a half (17 months). Among families interviewed after exiting early, their most frequently reported reason for leaving was moving. This is consistent with research demonstrating high rates of mobility among low-income families—the target population of Early Head Start programs.

- Most of the variability in participation is due to family not program or staff characteristics, although specific family characteristics included in the models are not strongly predictive of participation. We do not yet have a clear profile of families who are likely to leave early or participate at low levels that can help programs identify them early on and provide additional supports to their participation. This implies that programs that want to involve families need to be individualized in their approach, monitor participation, and find ways to proactively provide support and encourage continued participation when a family seems to be disengaging.

Participation and Service Quality Have Limited Associations to Outcomes

Although we can measure developmental outcomes, we are still lacking understanding of the factors that predict these outcomes. What causes children to develop as they do? What factors contribute the most to positive outcomes for low-income children, particularly for those being served in a program like Early Head Start? Baby FACES data highlight the various challenges associated with predicting child and family outcomes.

- Models that predicted child and family outcomes at age 3 using Early Head Start experiences (including participation) show that staff ratings of family involvement have the most predictive value of all the program participation constructs. Specifically, higher family involvement ratings are associated with fewer staff-reported behavior problems and higher competence. We would benefit from a deeper understanding of what these family involvement ratings represent—in its current form the single-item rating possibly represents a staff member's clinical judgment about a family. Future work to develop a more elaborated rating scale could help us understand whether this overall rating masks nuances that have stronger or weaker associations with child and family outcomes.

- Surprisingly, home visit service take-up is unrelated to child or family outcomes at age 3, and center take-up is only weakly related to children's auditory comprehension at age 3. These findings about services add to evidence from other large-scale studies that find similar results. Increasingly, the field of implementation science and practice is moving toward the idea that a one-size approach to offering services may not be the best strategy and that tailoring services to family needs is required to influence outcomes. Future research that systematically varies the amount of services that a family is offered may help identify what level of services is best for different families. Further, the quality of home visiting services children receive during their experience in Early Head Start was unrelated to outcomes (except for Spanish auditory comprehension). Center quality was modestly related to receptive vocabulary at age 3. It is not clear whether a constrained range of scores, particularly at the higher end of the home visit and center classroom quality scales, affected associations between quality and outcomes. Training and technical assistance

that focus on quality improvement, both on the lower and mid-higher end of the scale, could help raise quality overall, and could provide a wider distribution of scores, particularly in the high or excellent range and potentially allow a more nuanced understanding of how quality relates to outcomes.

Program Implementation Is Difficult to Measure

Program implementation continues to be an important but challenging feature to measure. Baby FACES drew on the experience of previous studies of Early Head Start, moving from program director self ratings on forms resembling those in the EHSREP to an interview format with scoring by researchers.

- The implementation analyses show that ratings are sensitive to the measurement and scoring approach. For example, we were not surprised to find that program self ratings are higher than researcher-scored implementation ratings. The cost of collecting detailed implementation data is significant and future studies will have to grapple with similar issues related to the trade-offs between using program self-report versus third party assessments and ratings.
- The current measures focus heavily on adherence to the performance standards, which set a threshold for expectations of implementation. Given the continuing maturation of the program and the field of implementation science, in future work Baby FACES may investigate additional areas beyond the performance standards which might be predictive of program outputs and outcomes and might be useful as a tool to guide program quality improvement efforts and deployment of technical assistance.

Study Limitations

Baby FACES is a descriptive study and thus there is no information on what would have happened to the children and families in the study in the absence of Early Head Start. Further, attrition from the program, or looking at it another way, self-selection of families to remain in the program, may limit the generalizability of the findings. Still, given the rich data collected over time, we attempt to assess more complex relationships between family, staff, and program characteristics and program participation, between staff and program characteristics and service quality, and between program experiences and outcomes. Overall, findings in general are in line with other studies of similar populations (see for example, EHSREP, ECLS-B, and FACES), but again, any associations cannot be interpreted as causal given the design of the study. They may, however, provide indications for future research and program improvement efforts.

Looking Ahead

Baby FACES has provided a comprehensive, longitudinal descriptive look at Early Head Start program, staff, services, and families, and provided opportunities to better understand the relationships among these different characteristics. Despite the questions that have been answered, there are many more that have been raised by the study and these have implications for future research. Some of the possibilities include:

- How do we measure and assess what makes a teacher, home visitor, and program effective?
- How can we measure implementation in ways that can inform program improvement?

- How can we disentangle the family involvement rating and understand the core elements needed to engage less involved families?
- What are the features of high quality home visits and how can we measure them? What makes a skilled home visitor?
- How does experience in high quality early care and education services influence children's development and growth over time? Are there critical thresholds of quality and amount of services that are required to enhance development?
- What are the reasons that families leave Early Head Start before children are 3 years old? Is it because they have achieved their goals and no longer need services, or do they leave for other reasons?

We are hopeful that the findings are useful for programs at the local level and for planning, training, policy-making, and technical assistance at the federal level.

REFERENCES

Abidin, R.R. *Parenting Stress Index, Third Edition: Professional Manual.* Odessa, FL: Psychological Assessment Resources, Inc., 1995.

Abidin, R.R., and T.R. Konold. *Parenting Alliance Measure Professional Manual.* Lutz, FL: Psychological Assessment Resources, Inc., 1999.

Achenbach T.M. and L.A. Rescorla. *Manual for the ASEBA Preschool Forms and Profiles.* Burlington, VT: University of Vermont Department of Psychiatry, 2000.

Achenbach T.M. and C.S. Edelbrock, Behavioral Problems and Competencies Reported by Parents of Normal and Disturbed Children Aged Four Through Sixteen. Monographs of the Society for Research in Child Development, vol. 46, no. 1, 1981, pp. 1-82.

Achenbach, Thomas M., and C.S. Edelbrock. *Manual for the Child Behavior Checklist and Revised Child Behavior Profile.* Burlington, VT: University of Vermont, Department of Psychology, 1983.

Administration for Children and Families. "Statement of the Advisory Committee on Services for Families with Infants and Toddlers." Washington, DC: U.S. Department of Health and Human Services, 1994. Available at www.bmcc.edu/Headstart/Advse.Commttee/index.html. Accessed December 7, 2009.

Administration for Children and Families. "Head Start Program Performance Standards." Final rule cited in *Federal Register,* November 5, 1996. Available at www.acf.hhs.gov/programs/hsb/performance. Accessed January 18, 2010.

Administration on Children, Youth and Families. "Building Their Futures: How Early Head Start Programs Are Enhancing the Lives of Infants and Toddlers in Low-Income Families." Washington, DC: U.S. Department of Health and Human Services, June 2001a.

Administration for Children and Families. "Head Start FACES: Longitudinal Findings on Program Performance: Third Progress Report." Washington, DC: U.S. Department of Health and Human Services, 2001b.

Administration for Children and Families. "Making a Difference in the Lives of Infants and Toddlers and Their Families: The Impacts of Early Head Start." Washington, DC: U.S. Department of Health and Human Services, June 2002a.

Administration for Children and Families. *Pathways to Quality and Full Implementation in Early Head Start Programs.* Washington, DC: U.S. Department of Health and Human Services, June 2002b.

Administration for Children and Families. *The Role of Early Head Start Programs in Addressing the Child Care Needs of Low-Income Families with Infants and Toddlers: Influences on Child Care Use and Quality*. Washington, DC: U.S. Department of Health and Human Services, 2004.

Administration for Children and Families. "Head Start Performance Measures Center Family and Child Experiences Survey (FACES 2000) Technical Report." Washington, DC: U.S. Department of Health and Human Services, 2006.

Administration for Children and Families. "Program Information Report User's Guide." Washington, DC: U.S. Department of Health and Human Services, 2007.

Administration for Children and Families. "FACES 2003 Fifth Progress Report: Technical Appendix." Washington, DC: U.S. Department of Health and Human Services, 2008.

Administration for Children and Families. "Head Start Program Performance Standards." Cited in *Federal Register,* October 1, 2009a. Available at http://eclkc.ohs.acf.hhs.gov/hslc/standards/Head%20Start%20Requirements/45%20CFR%20Chapter%20XIII/45%20CFR%20Chap%20XIII_ENG.pdf. Accessed February 7, 2012.

Administration for Children and Families. "How the Performance Standards Support New Early Head Start Programs: Lessons Learned from Research." Washington, DC: U.S. Department of Health and Human Services, 2009b.

Administration for Children and Families. *Justification of Estimates for Appropriations Committees. Fiscal Year 2011*. Washington, DC: ACF, 2010. Available at http://www.acf.hhs.gov/programs/olab/budget/2011/2011_all.pdf. Accessed June 1, 2010.

Aikens, N., L. Tarullo, L. Hulsey, C. Ross, J. West, and Y. Xue. "ACF-OPRE Report: A Year in Head Start: Children, Families, and Programs." Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation, 2010.

Anderson, S.E., and R.C. Whittaker. "Household Routines and Obesity in U.S. Preschool-Aged Children." *Pediatrics,* vol. 125, no. 3, March 2010, pp. 420–428.

Andreassen, C., and P. Fletcher. *Early Childhood Longitudinal Study, Birth Cohort (ECLS–B) Psychometric Report for the 2-Year Data Collection.* (NCES 2007–084). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2007.

Appleyard, K., B. Egeland, M.H.M. van Dulmen, and L.A. Sroufe. "When More Is Not Better: The Role of Cumulative Risk in Child Behavior Outcomes." *Journal of Child Psychology and Psychiatry*, vol. 46, no. 3, 2005, pp. 235–245.

Astuto, J., and L. Allen. "Home Visitation and Young Children: An Approach Worth Investing In?" *Social Policy Report*, vol. 23, no. 4, 2009.

Bagnato, S.J., H.K. Suen, and A.V. Fevola. " 'Dosage' Effects on Developmental Progress During Early Childhood Intervention: Accessible Metrics for Real-Life Research and Advocacy." *Infancy and Young Children,* vol. 24, no. 2, 2011, pp. 117–132.

Barlow, S.E., and W.H. Dietz. "Obesity Evaluation and Treatment: Expert Committee Recommendations." *Pediatrics,* vol. 102, no. 3, 1998, p. e29. Available at http://www.pediatrics.org/cgi/content/full/102/3/e29. Accessed March 7, 2011.

Bayley, N. *Bayley Scales of Infant Development—Second Edition.* San Antonio, TX: Psychological Corporation, 1993.

Bayley, N. *Bayley Scales of Infant Development-III*. San Antonio, TX: Psychological Corporation, 2006.

Berkel, C., A.M. Mauricio, E. Schoenfelder, and I.N. Sandler. "Putting the Pieces Together: An Integrated Model of Program Implementation." *Prevention Science,* vol. 12, 2010, pp. 23–33.

Bisceglia, R., M. Perlman, D. Schaack, and J. Jenkins. "Examining the Psychometric Properties of the Infant-Toddler Environment Rating Scale-Revised Edition in a High-Stakes Context." *Early Childhood Research,* vol. 24, 2009, pp. 121–132.

Bogard, K., F. Traylor, and R. Takanishi. "Teacher Education and PK Outcomes: Are We Asking the Right Questions?" *Early Childhood Research Quarterly*, vol. 23, no. 1, 2008, pp. 1–6.

Boller, K., C. Vogel, R. Cohen, N. Aikens, and K. Hallgren. "Home Visit Characteristics and Content Form." Princeton, NJ: Mathematica Policy Research, 2009.

Bradley, R., B. Caldwell, and F. Corwyn. "The Child Care HOME Inventories: Assessing the Quality of Family Child Care Homes." *Early Childhood Research Quarterly,* vol. 18, no. 3, 2003, pp. 294–309.

Bradley, R.H. "The Home Environment." In *The Crisis in Youth Mental Health: Critical Issues and Effective Programs Volume 4: Early Intervention Programs and Policies*, edited by H.E. Fitzgerald, R. Zucker, and K. Freeark. Westport, CT: Praeger Publishers/Greenwood Publishing Group, 2006.

Bradley, R.H., and R.F. Corwyn. "Externalizing Problems in Fifth Grade: Relations with Productive Activity, Maternal Sensitivity, and Harsh Parenting from Infancy Through Middle Childhood." *Developmental Psychology,* vol. 43, no. 6, 2007, pp. 1390–1401.

Bradley, R.H., R. Corwyn, M. Burchinal, H.P. McAdoo, and G. Coll. "The Home Environments of Children in the United States Part II: Relations with Behavioral Development Through Age Thirteen." *Child Development,* vol. 72, 2001, pp. 1868–1886.

Brady-Smith, C., R.C. Fauth, and J. Brooks-Gunn. *Background and Psychometric Information for the Child-Parent Interaction Rating Scales for the Three-Bag Assessment: 14-, 24-, and 36-Month Waves.* New York: Columbia University, Teachers College, 2005.

Brady-Smith, C., C. O'Brien, L. Berlin, and A. Ware. *24-Month Child–Parent Interaction Rating Scales for the Three-Bag Assessment.* New York: Columbia University, Teachers College, 1999.

Brady-Smith, C., C. O'Brien, L. Berlin, A. Ware, and R. Fauth. *36-Month Child–Parent Interaction Rating Scales for the Three-Bag Assessment. Early Head Start Research and Evaluation Project.* Unpublished scales. New York: Columbia University, Teachers College, National Center for Children and Families, 2000.

Briggs-Gowan, M.J. *Brief Infant-Toddler Social and Emotional Assessment Examiner's Manual.* San Antonio, TX: Harcourt Assessment, 2006.

Briggs-Gowan, M.J., and A.S. Carter. *The Brief Infant–Toddler Social and Emotional Assessment (BITSEA).* San Antonio, TX: Harcourt Assessment, 2006.

Briggs-Gowan, M.J., A.S. Carter, J.R. Irwin, K. Wachtel, and D.V. Cicchetti. "The Brief Infant-Toddler Social and Emotional Assessment: Screening for Social-Emotional Problems and Delays in Competence." *Journal of Pediatric Psychology,* vol. 29, no. 2, 2004, pp. 143–155.

Brookes, S.J., J.M. Ispa, J.A. Summers, K.R. Thornburg, and V. Lane. "Building Successful Home Visitor-Mother Relationships and Reaching Program Goals: A Qualitative Look at Contributing Factors." *Early Childhood Research Quarterly,* vol. 21, no. 1, 2006, pp. 25–45.

Burchinal, M., K. Kainz, K. Cai, K. Tout, M. Zaslow, I. Martinez-Beck, and C. Rathgeb. "Child Care Quality and Child Outcome: Multiple-Studies Analyses." Presentation at the Working Meeting of the Developing the Next Wave of Quality Measures for Early Childhood and School-Age Programs, January 2008a.

Burchinal, M., C. Howes, R. Pianta, D. Bryant, D. Early, R. Clifford, and O. Barbarin. "Predicting Child Outcomes at the End of Kindergarten from the Quality of Pre-Kindergarten Teacher-Child Interactions and Instruction." *Applied Developmental Science*, vol. 12, no. 3, 2008b, pp. 140–153.

Burchinal, M., K. Kainz, K. Cai, K. Tout, M. Zaslow, I. Martinez-Beck, and C. Rathgeb. "Early Care and Education Quality and Child Outcomes." Publication no. 2009-15. Washington, DC: Administration for Children and Families, Office of Planning, Research and Evaluation, May 2009.

Burchinal, M.R., J.E. Roberts, L.A. Nabors, and D.M. Bryant. "Quality of Center Child Care and Infant Cognitive and Language Development." *Child Development,* vol. 67, 1996, pp. 606–620.

Burchinal, M.R., J.E. Roberts, R. Riggins, S.A. Zeisel, E. Neebe, and D. Bryant. "Relating Quality of Center-Based Child Care to Early Cognitive and Language Development Longitudinally." *Child Development*, vol. 71, 2000, pp. 339–357.

Caldwell B. and Bradley, R. . *Home Observation for Measurement of the Environment (HOME)-Revised Edition.* University of Arkansas, Little Rock, 1984.

Caldwell, B.M., and R.H. Bradley. *Home Observation for Measurement of the Environment: Administration Manual.* Tempe, AZ: Arizona State University, Family and Human Dynamics Research Institute, 2003.

Caronongan, P., E. Moiduddin, J. West, and C.A. Vogel. *Children in Early Head Start and Head Start: A Profile of Early Leavers. Baby FACES and FACES 2009 Research Brief.* OPRE report no. 2014-54. Washington, DC. U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation, 2014.

Carta, J.J., C.R. Greenwood, D. Walker, R. Kaminski, R. Good, S. McConnell, and M. McEvoy. "Individual Growth and Development Indicators (IGDIs): Assessment That Guides Intervention for Young Children." *Young Exceptional Children Monograph Series,* vol. 4, 2002, pp. 15–28.

Carta, J., C. Greenwood, G. Luze, G. Cline, and S. Kuntz. "Developing a General Outcome Measure of Growth in Social Skills for Infants and Toddlers." *Journal of Early Intervention*, vol. 26, no. 2, 2004, pp. 91-114.

Carta, J.J., C.R. Greenwood, D. Walker, and J. Buzhardt. Using IGDIs: Monitoring Progress and Improving Intervention for Infants and Young Children. Baltimore, MD: Paul H. Brookes, 2010.

Carter, A.S., and M.J. Briggs-Gowan. Infant–Toddler Social and Emotional Assessment–Revised (ITSEA), San Antonio, TX: Pearson, 2000.

Caughy, M.O.B., J.A. DiPietro, and D.M. Strobino. "Day-Care Participation as a Protective Factor in the Cognitive Development of Low-Income Children." Child Development, vol. 65, no. 2, 1994, pp. 457–471.

Center for Human Resource Research. *NLSY79 Child and Young Adult Data Users Guide: A Guide to the 1986–2006 Child Data 1994–2006 Young Adult Data.* Columbus, OH: Ohio State University, CHRR, 2009. Available at http://www.nlsinfo.org/pub/usersvc/Child-Young-Adult/2006ChildYA-DataUsersGuide.pdf.

Chazan-Cohen, Rachel, Ellen Kisker, Jill Constantine, and Helen Raikes. "What About Dosage? Level of Participation in Early Head Start Services and Child and Parent Outcomes." Poster presented at the Society for Research in Child Development Biennial Meeting, Denver, CO, April 2009.

Clarke-Stewart, K.A. "Infant Day Care: Maligned or Malignant?" *American Psychologist,* vol. 44, no. 2, 1989, pp. 266–273.

Cook, G., and L. Roggman. *Parenting Interactions with Children: Checklist of Observations Linked to Outcomes (PICCOLO); Technical Report.* Logan, UT: Utah State University, Early Intervention Research Institute, 2009.

Cox, M.J. *Qualitative Ratings for Parent–Child Interaction at 24–36 Months of Age.* Unpublished manuscript. Chapel Hill, NC: University of North Carolina at Chapel Hill, Department of Psychology, 1997.

Cox, J.L., J.M. Holden, and R. Sagovsky. "Detection of Postnatal Depression: Development of the 10-Item Edinburgh Postnatal Depression Scale." *British Journal of Psychiatry,* vol. 150, 1987, pp. 782–786.

Cronbach, L.J. "Coefficient Alpha and the Internal Structure of Tests." *Psychometrika,* vol. 16, no. 3, 1951, pp. 297–334.

Crosnoe, Robert. "The Diverse Experiences of Hispanic Students in the American Educational System." *Sociological Forum,* vol. 20, 2005, pp. 561–588.

Cummings, E.M., and P. Davies. *Children and Marital Conflict: The Impact of Family Dispute and Resolution.* New York: Guilford Press, 1994.

Damschroder, L.J., and H.J. Hagedorn. "A Guiding Framework and Approach for Implementation Research in Substance Use Disorders Treatment." *Psychology of Addictive Behaviors,* vol. 25, no. 2, 2011, pp. 194–205.

Del Grosso, P., R. Kleinman, A. Esposito, E. Martin, and D. Paulsell. "Assessing the Evidence of Effectiveness of Home Visiting Program Models Implemented in Tribal Communities." Princeton, NJ: Mathematica Policy Research, February 2011.

Deparle, J., and R. Gebeloff. "Food Stamp Use Soars, and Stigma Fades." *New York Times,* November 28, 2009. Available at http://www.nytimes.com/2009/11/29 /us/29food stamps.html?_r=1. Accessed January 12, 2010.

de Schipper, E. J., M. Riksen-Walraven, and S.A.E. Geurts. "Effects of child-caregiver ratio on interactions between caregivers and children in child-care centers: An experimental study." *Child Development*, vol. 77, no. 4, 2006, 861-874.

Dunn, L.M., and D.M. Dunn. *Peabody Picture Vocabulary Test—Fourth Edition.* Bloomington, MN: NCS Pearson, Inc., 2007.

Dunst, Carol J., Carol M. Trivette, and Angela Deal (eds.). *Supporting and Strengthening Families: Methods, Strategies, and Practices.* Newton, MA: Brookline Books, 1994.

Durlak, Joseph A., and Emily P. DuPre. "Implementation Matters: A Review of Research on the Influence of Implementation on Program Outcomes and the Factors Affecting Implementation." *American Journal of Community Psychology,* vol. 41, nos. 3–4, June 2008, pp. 327–350.

Early, D.M., D.M. Bryant, R.C. Pianta, R.M. Clifford, M.R. Burchinal, S. Ritchie, C. Howes, and O. Barbarin. "Are Teachers' Education, Major, and Credentials Related to Classroom Quality and Children's Academic Gains in Pre-Kindergarten?" *Early Childhood Research Quarterly*, vol. 21, 2006, pp. 174–195.

Early Head Start National Resource Center at ZERO TO THREE. Available at http://www.ehsnrc.org/faqs.htm. Accessed March 8, 2010.

Early Head Start Tip Sheet. "Continuity of Care in Early Head Start Group Settings." Available at http://www.ehsnrc.org/Publications/English%20Tip%20Sheets/TIP_SHEET_21.PDF. Accessed April 25, 2011.

Egeland, B., and T. Kreutzer. "A Longitudinal Study of the Effects of Maternal Stress and Protective Factors on the Development of High-Risk Children." In *Life-Span Developmental Psychology: Perspectives on Stress and Coping,* edited by E.M. Cummings, A.L. Greene, and K.H. Karraker. Hillsdale, NJ: Lawrence Erlbaum Associates, 1991.

Elicker, J., I. Noppe, L. Noppe, and C. Fortner-Wood. "The Parent-Caregiver Relationship Scale: Rounding Out the Relationship System in Infant Care." *Early Education and Development,* vol. 8, no. 1, 1997, pp. 83–100.

Espinosa, L.M. "Social, Cultural, and Linguistic Features of School Readiness in Young Latino Children." *School Readiness and Social-Emotional Development: Perspectives on Cultural Diversity*, edited by B. Bowman and K. Moore. Washington, DC: National Black Child Development Institute, Inc., 2006.

Federal Interagency Forum on Child and Family Statistics. *America's Children: Key National Indicators of Well-Being, 2009*. Washington, DC: U.S. Government Printing Office, 2009.

Fenson, L., S. Pethick, C. Renda, J.L. Cox, P.S. Dale, and J.S. Reznick. "Short-Form Versions of the MacArthur Communicative Development Inventories." *Applied Psycholinguistics,* vol. 21, 2000, pp. 95–115.

Fenson, L., V.A. Marchman, D.J. Thal, P.S. Dale, J.S. Reznick, and E. Bates. *MacArthur-Bates Communicative Development Inventories: User's Guide and Technical Manual.* Baltimore, MD: Brookes, 2007.

Fernández, M.C., B.Z. Peason, V.M. Umbel, and M. Molinet-Molina. "Bilingual Receptive Vocabulary in Hispanic Preschool Children." *Hispanic Journal of Behavioral Sciences,* vol. 14, no. 2, 1992, pp. 268–276.

Flanagan, K.D., and J. West. *Children Born in 2001: First Results from the Base Year of the Early Childhood Longitudinal Study, Birth Cohort (ECLS–B).* (NCES 2005–036). Washington, DC: National Center for Education Statistics, 2004.

Fuller, Bruce, Edward Bein, Margaret Bridges, Neal Halfon, Sunyoung Jung, Sophia Rabe-Hesketh, and Alice Kuo. "Maternal Practices That Influence Hispanic Infants' Health and Cognitive Growth." *Pediatrics,* vol. 125, no. 2, February 2010, pp. e324–e332.

Fuligni, A., and J. Brooks-Gunn. "Mother-child interactions in Early Head Start: Age and ethnic differences in low-income dyads." *Parenting: Science and Practice*, vol. 13, 2013, pp. 1–26.

Gabe, Thomas. "Poverty in the United States: 2008." Washington, DC: Congressional Research Service, 2009.

Gerber, Emily B., Marcy Whitebook, and Rhona S. Weinstein. "At the Heart of Child Care: Predictors of Teacher Sensitivity in Center-Based Child Care." *Early Childhood Research Quarterly*, vol. 22, no. 3, 2007, pp. 327–346.

Greenwood, C. R., J.J. Carta, & D. Walker. "Individual growth and development indicators (IGDIs): Tools for assessing intervention results for infants and toddlers." In B. Heward et al. (Eds.), *Focus on Behavior Analysis in Education: Achievements, Challenges, and Opportunities.* Columbus, OH: Pearson/Prentice-Hall, 2005, pp.103-124.

Greenwood, C.R., J.J. Carta, D. Walker, K. Hughes, and M. Weathers. "Preliminary Investigations of the Application of the Early Communication Indicator (ECI) for Infants and Toddlers." *Journal of Early Intervention,* vol. 28, no. 3, 2006, pp. 178–196.

Greenwood, C.R., G.J. Luze, G. Cline, S. Kuntz, and C. Leitschuh. "Developing a General Outcome Measure of Growth in Movement for Infants and Toddlers." *Topics in Early Childhood Special Education*, vol. 22, no.3, 2002, pp. 143-157.

Greenwood, C.R., D. Walker, and J. Buzhardt. "The Early Communication Indicator (ECI) for Infants and Toddlers: Growth Norms from Two States." *Journal of Early Intervention,* vol. 32, 2010, pp. 310–334.

Hallgren, K., N. Aikens, K. Boller, J. Stein, L.A. Roggman, G.A. Cook, and M.S. Innocenti. "Home Visit Rating Scales-Adapted (HOVRS-A) Manual." Princeton, NJ: Mathematica Policy Research, 2009.

Hamre, Bridget K. and Robert C. Pianta. "Self-Reported Depression in Nonfamilial Caregivers: Prevalence and Associations with Caregiver Behavior in Child-Care Settings." *Early Childhood Research Quarterly*, vol. 19, no. 2, 2004, pp. 297–318.

Harms, T., D. Cryer, and R.M. Clifford. *Infant/Toddler Environment Rating Scale.* New York: Columbia University, Teachers College Press, 1990.

Harms, T., D. Cryer, and R.M. Clifford. *Infant/Toddler Environment Rating Scale: Revised Edition.* New York: Columbia University, Teachers College Press, 2003.

Harrison, P.L., & Oakland, T. (2003), *Adaptive Behavior Assessment System – Second Edition.* San Antonio, TX: Psychological Corporation, 2003.

Hayes, C.D., J.L. Palmer, and M.J. Zaslow. *Who Cares for America's Children? Child Care Policy for the 1990s*. Washington, DC: American Family Press, 1990.

Hestenes, L.L., D.J. Cassidy, A.V. Hegde, and J.X. Lower. "Quality in Inclusive and Noninclusive Infant and Toddler Classrooms." *Journal of Research in Childhood Education*, vol. 22, no. 1, 2007, pp. 69–84.

Hillemeier, M.M., E.M. Foster, B. Heinrichs, B. Heier, and Conduct Problems Prevention Research Group. "Racial Differences in Parental Reports of Attention-Deficit/Hyperactivity Disorder Behaviors." *Journal of Developmental and Behavioral Pediatrics,* vol. 28, no. 5, 2007, pp. 353–361.

Ispa, J.M., M.A. Fine, L.C. Halgunseth, S. Harper, J. Robinson, L. Boyce, and C. Brady-Smith, C. "Maternal intrusiveness, maternal warmth, and mother-toddler relationship outcomes: Variations across low-income ethnic and acculturation groups." *Child Development*, vol. 75, 2004, pp. 1613–1631.

Jackson-Maldonado, D., D. Thal, V. Marchman, T. Newton, L. Fenson, and B. Conboy. *MacArthur Inventarios del Desarrollo de Habilidades Comunicativas. User's Guide and Technical Manual*. Baltimore, MD: Brookes, 2003.

Kahn, R.S., D. Brandt, and R.C. Whitaker. "Combined Effect of Mothers' and Fathers' Mental Health Symptoms on Children's Behavioral and Emotional Well-Being." *Archives in Pediatric and Adolescent Medicine,* vol. 158, no. 8, 2004, pp. 721–729.

Korfmacher, J., B. Green, M. Spellmann, and K.R. Thornburg. "The Helping Relationship and Program Participation in Early Childhood Home Visiting." *Infant Mental Health Journal,* vol. 28, no. 5, 2007, pp. 459–480.

Korfmacher, J., B. Green, F. Staerkel, C. Peterson, G. Cook, L. Roggman, R.A. Faldowski, and R. Schiffman. "Parent Involvement in Early Childhood Home Visiting." *Child Youth Care Forum,* vol. 37, 2008, pp. 171–196.

Kuczmarski, R.J., C.L. Ogden, S.S. Guo, L.M. Grummer-Strawn, K.M. Flegal, Z. Mei, R. Wei, L.R. Curtin, A.F. Rocje, and C.L. Johnson. "2000 CDC Growth Charts for the United States: Methods and Development." Vital Health Statistics, vol. 11, no. 246, 2002, pp. 1–190.

Lamb, M.E. "Nonparental child care: Context, quality, correlates, and consequences." In W. Damon, I.E. Sigel, & K.A. Renninger (Eds.), *Handbook of child psychology: Child psychology in practice,* vol. 4, 5th ed., pp. 73-134. New York: Wiley, 1998.

La Paro, K.M., B.K. Hamre, and R.C. Pianta. *Classroom Assessment Scoring System Manual, Toddler (CLASS-T)*. Baltimore, MD: Paul H. Brookes Publishing Co., 2012.

La Paro, K.M., A.C. Williamson, and B. Hatfield. "Assessing Quality in Toddler Classrooms Using the CLASS-Toddler and the ITERS-R." *Early Education and Development*, vol. 25, no. 6, 2014, pp. 875–893.

Love, J.M., E.E. Kisker, C. Ross, H. Raikes, J. Constantine, K. Boller, J. Brooks-Gunn, R. Chazan-Cohen, L. Banks Tarullo, C. Brady-Smith, A.S. Fuligni, P.Z. Schochet, D. Paulsell, and C. Vogel. "The Effectiveness of Early Head Start for 3-Year-Old Children and Their Parents: Lessons for Policy and Programs." *Developmental Psychology,* vol. 41, no. 6, 2005, pp. 885–901.

Love, John M., Rachel Chazan-Cohen, Helen Raikes, and Jeanne Brooks-Gunn (eds.). "What Makes a Difference? Early Head Start Evaluation Findings in a Longitudinal Context." *Monograph of the Society for Research in Child Development,* vol. 78, no. 1, 2013.

Luze, G.J., D.L. Linebarger, C.R. Greenwood, J.J. Carta, D. Walker, C. Leitschuh, and J.B. Atwater. "Developing a General Outcome Measure of Growth in the Expressive Communication of Infants and Toddlers." *School Psychology Review*, vol. 30, no. 3, 2001, pp. 383–406.

Mashburn, A.J., R.C. Pianta, B.K. Hamre, J.T. Downer, O.A. Barbarin, D. Bryant, M. Burchinal, D.M. Early, and C. Howes. "Measures of Classroom Quality in Prekindergarten and Children's Development of Academic, Language, and Social Skills." *Child Development,* vol. 79, no. 3, 2008, pp. 732–749.

Mathematica Policy Research. "Baby FACES and Building Strong Families: Video Coding Manual for the Two-Bag Task." Unpublished scales adapted from Brady-Smith et al. *36-Month Child-Parent Interaction Rating Scales for the Three-Bag Assessment. Early Head Start Research and Evaluation Project,* 2000; and Cox. *Qualitative Ratings for Parent-Child Interaction at 24–36 Months of Age,* 1997, 2010.

Mathematica Policy Research Early Learning Initiative Evaluation Team. "Home Visit Characteristics and Content Form." Princeton, NJ: Mathematica Policy Research, 2009.

Moos, R.H., and B.S. Moos. *Family Environment Scale*. Redwood City, CA: Mind Garden, 2002.

National Bureau of Labor Statistics. "Current Unemployment Rate." In *Current Population Survey*. Washington, DC: NBLS, 2009. Available at http://www.bls.gov/bls/unemployment.htm. Accessed February 8, 2010.

National Child Care Information and Technical Assistance Center. "QRS Planning Tool: Standards for Early Childhood Education Programs, May 2008." Available at http://nccic.acf.hhs.gov/poptopics/qrs-pt-ece.html. Accessed July 30, 2009.

National Institute of Child Health and Human Development Early Child Care Research Network. "Early Child Care and Children's Development Prior to School Entry: Results from the NICHD Study of Early Child Care." Rockville, MD: NICHD, 1996.

National Institute of Child Health and Human Development Early Child Care Research Network. "Early Child Care and Children's Development Prior to School Entry: Results from the NICHD Study of Early Child Care." Rockville, MD: NICHD, 1997.

National Institute of Child Health and Human Development Early Child Care Research Network. "Early Child Care and Children's Development Prior to School Entry: Results from the NICHD Study of Early Child Care." Rockville, MD: NICHD, 1998.

National Institute of Child Health and Human Development Early Child Care Research Network. "Child Care and Mother-Child Interaction in the First Three Years of Life." *Developmental Psychology,* vol. 35, 1999, pp. 1399–1413.

National Institute of Child Health and Human Development Early Child Care Research Network. "Early Child Care and Children's Development Prior to School Entry: Results from the NICHD Study of Early Child Care." Rockville, MD: NICHD, 2002.

National Institute of Child Health and Human Development Early Child Care Research Network. "Early Child Care and Children's Development Prior to School Entry: Results from the NICHD Study of Early Child Care." Rockville, MD: NICHD, 2006.

National Research Council and Institute of Medicine. *From Neurons to Neighborhoods: The Science of Early Childhood Development,* edited by J.P. Shonkoff and D. Phillips. Washington, DC: National Academies Press, 2000.

Newborg, J., J.R. Stock, L. Wnek, J. Guidubaldi. & J. Svinicki. *The Battelle Developmental Inventory*. Allen, TX: DLM/Teaching Resources, 1984.Newborg, J. *Battelle Developmental Inventory, Second Edition.* Itasca, IL: Riverside, 2004.

Oller, D.K., and R.E. Eilers (eds.). *Language and Literacy in Bilingual Children*. Tonawanda, NY: Multilingual Matters, 2002.

Parlakian, R., and N.L. Seibel. *Building Strong Foundations: Practical Guidance for Promoting the Social-Emotional Development of Infants and Toddlers*. Washington, DC: Zero to Three, 2002.

Paulsell, Diane, Kimberly Boller, Kristin Hallgren, and Andrea Mraz Esposito. "Assessing Home Visiting Quality: Dosage, Content, and Relationships." Washington, DC: Zero to Three, July 2010.

Peterson, C. A., & Roggman, L. A. Home visiting interventions: Challenges faced in learning about what really happens." Conference on Research Issues in Early Intervention, San Diego, CA., February 2006

Phillips, D., D. Mekos, S. Scarr, K. McCartney, and M. Abbott-Shim. "Within and Beyond the Classroom Door: Assessing Quality in Child Care Centers." *Early Childhood Research Quarterly*, vol. 15, 2000, pp. 475–496.

Phillipsen, L.C., M.R. Burchinal, C. Howes, and D. Cryer. "The Prediction of Process Quality from Structural Features of Child Care." *Early Childhood Research Quarterly*, vol. 12, 1997, pp. 281–303.

Pianta, R.C., K.M. La Paro, and B.K. Hamre. *Classroom Assessment Scoring System (CLASS). Pre-K Version*. Baltimore, MD: Paul H. Brookes Publishing Co., 2008.

Pianta, R.C., K.M. La Paro, and B.K. Hamre. *Classroom Assessment Scoring System Pilot Toddler Manual (CLASS-T)*. Charlottesville, VA: Teachstone, Inc., 2010.

Radloff, L.S. "The CES-D Scale: A Self-Report Depression Scale for Research in the General Population." *Applied Psychological Measurement,* vol. 1, 1977, pp. 385–401.

Raikes, H., B.L. Green, J. Atwater, E. Kisker, J. Constantine, and R. Chazan-Cohen. "Involvement in Early Head Start Home Visiting Services: Demographic Predictors and Relations to Child and Parent Outcomes." *Early Childhood Research Quarterly,* vol. 21, no. 1, 2006, pp. 2–24.

Rathbun, A., and J. West. *From Kindergarten Through Third Grade: Children's Beginning School Experiences.* (NCES 2004-007). Washington, DC: National Center for Education Statistics, 2004.

Raudenbush, S.W., and A.S. Bryk. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage, 2002.

Roggman, L., G. Cook, C. Peterson, and H. Raikes. "Who Drops Out of Early Head Start Home Visiting Programs?" *Early Education and Development*, vol. 19, no. 4, 2008a, pp. 574–599.

Roggman, L., G. Boyce, and M. Innocenti. *Developmental Parenting: A Guide for Early Childhood Practitioners*. Baltimore, MD: Paul H. Brooks, 2008b.

Roggman, L.A., K. Christiansen, G.A. Cook, V.K. Jump, L.K. Boyce, and C.A. Peterson. "Home Visits: Measuring How They Work." Paper presented at the Early Intervention Research Institute Mini-Conference, Logan, UT, 2006a.

Roggman, L.A., G.A. Cook, V.K. Jump Norman, K. Christiansen, L.K. Boyce, M.S. Innocenti, N. Aikens, K. Boller, D. Paulsell, and K. Hallgren. *Home Visit Rating Scales-Adapted (HOVRS-A)*. Unpublished document. Princeton, NJ: Mathematica Policy Research, 2009.

Roggman, Lori, Gina Cook, Vonda Jump, Lisa Boyce, and Mark Innocenti. "Home Visit Rating Scales (HOVRS)." Unpublished manuscript. Logan, UT: Utah State University, 2006b.

Ross, C., E. Moiduddin, C. Meagher, and B. Carlson. "The Chicago Program Evaluation Project: A Picture of Early Childhood Programs, Teachers, and Preschool-Age Children in Chicago." Princeton, NJ: Mathematica Policy Research, 2008.

Ross, C.E., J. Mirowsky, and J. Huber. "Dividing Work, Sharing Work, and in Between: Marriage Patterns and Depression." *American Sociological Review,* vol. 48, 1983, pp. 809–823.

Ruopp, R., J. Travers, F. Glantz, and C. Coelen. "Children at the Center." Cambridge, MA: Abt Associates, 1979.

Sameroff, A.J., R. Seifer, R. Barocas, M. Zax, and S. Greenspan. "Intelligence Quotient Scores of 4-Year-Old Children: Social Environmental Risk Factors." *Pediatrics*, vol. 79, 1987, pp. 343–350.

Sampson, Robert J. *Great American City: Chicago and the Enduring Neighborhood Effects*. Chicago: University of Chicago Press, 2012.

Scarr, S., M. Eisenberg, and K. Deater-Deckard. "Measurement of Quality in Child Care Centers." *Early Childhood Research Quarterly,* vol. 9, 1994, pp. 131–151.

Schaefer, E. "Goals for Parent and Future-Parent Education: Research on Parental Beliefs and Behavior." *Elementary School Journal,* vol. 91, no. 3, 1991, pp. 239–247.

Schaefer, E., and M. Edgerton. "Parental and Child Correlates of Parental Modernity." In *Parental Belief Systems: The Psychological Consequences for Children,* edited by I.E. Sigel. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 1985.

Schafer, J.L., and J. Kang. "Average Causal Effects from Nonrandomized Studies: A Practical Guide and Simulated Example." *Psychological Methods,* vol. 13, no. 4, 2008, p. 279.

Squires, J., D. Bricker, and E. Twombly. *The ASQ:SE User's Guide: for the Ages and Stages Questionnaires: Social Emotional.* Baltimore, MD: Paul H. Brookes, 2002.

Squires, J., E. Twombly, D. Bricker, and L. Potter. *ASQ-3 User's Guide*. Baltimore, MD: Paul H. Brookes, 2009.

Substance Abuse and Mental Health Services Administration. *Results from the 2008 National Survey on Drug Use and Health: National Findings*. HHS Publication no. SMA 09-4434. Rockville, MD: Office of Applied Studies, 2009.

Sweet, Monica A., and Mark I. Appelbaum. "Is Home Visiting an Effective Strategy? A Meta-Analytic Review of Home Visiting Programs for Families with Young Children." *Child Development,* vol. 75, no. 5, 2004, pp. 1435–1456.

Tamis-LeMonda, C.S., T.N. Cristofaro, E.T. Rodríguez, and M.H. Bornstein. "Early Language Development: Social Influences in the First Years of Life." In *Child Psychology: A Handbook of Contemporary Issues, 2*, edited by L. Balter and C.S. Tamis-LeMonda. Philadelphia: Taylor and Francis Group, 2005.

Tamis-LeMonda, Catherine, Jaqueline D. Shannon, Natasha Cabrera, and Michael E. Lamb. "Fathers and Mothers at Play with Their 2- and 3-Year-Olds: Contributions to Language and Cognitive Development." *Child Development*, vol. 75, 2004, pp. 1806-1820.

Tarullo, L., and C. Ross. "Charting the Relationship of Head Start Teacher Credentials and Classroom Quality Over Time." Presentation at the Society for Research in Child Development Biennial Meeting, Denver, CO, April 2009.

Thomason, A.C., and K.M. La Paro. "Measuring the Quality of Teacher-Child Interactions in Toddler Child Care." *Early Education and Development*, vol. 20, no. 2, 2009, pp. 285–304.

U.S. Census Bureau. "Survey of Income and Program Participation." Washington DC: Department of the Census, 1996. Available at http://www.census.gov/sipp/index.html. Accessed January 16, 2010.

U.S. Congress. "Improving Head Start Act, 2007." Conference report no. 110-439 (to accompany H.R. 1429). Available at http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=110_cong_reports&docid=f:hr439.110.pdf. Accessed November 9, 2007.

U.S. Department of Health and Human Services. "Qualifications for Early Head Start Infant and Toddler Center-Based Staff." ACF information memorandum 10-06. Available at [http://eclkc.ohs.acf.hhs.gov/hslc/Program%20Design%20and%20Management/Head%20Start%20Requirements/IMs/2010/resour_ime_006_102610.html]. Accessed November 10, 2010.

Vagh, S.B., J. Mançilla-Martinez, and B.A. Pan. "An Upward Extension Based on the Inventarios del Desarrollo de Habilidades Comunicativas Short Form for Use with Children of Ages 30-37 Months from Low-Income Bilingual Families." Unpublished manuscript.

Vandell, D. L., & Wolfe, B. (2000). "Child care quality: Does it matter and does it need to be improved?" Washington, D.C.: Office of the Assistant Secretary for Planning and Evaluation: U.S. Department of Health and Human Services. Available at [http://aspe.hhs.gov/hsp/ccquality00/execsum.htm#top]. Accessed September 24 2014.

Vogel, C., N. Aikens, A. Burwick, L. Hawkinson, A. Richardson, L. Mendenko, and R. Chazan-Cohen. "Findings from the Survey of Early Head Start Programs: Communities, Programs and Families." Washington, DC: U.S. Department of Health and Human Services, Children and Family Services, 2006.

Vogel, C., K. Boller, Y. Xue, R. Blair, N. Aikens, A. Burwick, and J. Stein. *Learning as We Go: A First Snapshot of Early Head Start Programs, Staff, Families, and Children.* OPRE Report no. 2011-7. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation, 2011.

Vogel, C.A., P. Caronongan, J. Thomas, E. Bandel, Y. Xue, J. Henke, and L. Murphy. *Toddlers in Early Head Start: A Portrait of 2-Year-Olds, Their Families, and the Programs Serving Them.* OPRE report no. 2015-10. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation, 2015.

Walker, D., B. Hart, D.L. Linebarger, and K.P. Parsley. Caregiver Communication Measure. Unpublished manuscript, Early Childhood Research Institute on Measuring Growth and Development, Juniper Gardens Children's Project, University of Kansas, Kansas City, KS, 1998.

Whitaker, J.E., and B.J. Harden. "Beyond ABCs and 123s: Enhancing Teacher-Child Relationship Quality to Promote Children's Behavioral Development." *National Head Start Association Dialog*, vol. 13, no. 3, 2010, pp. 185–191.

Whitebook, M., C. Howes, and D. Phillips. "Who Cares? Child Care Teachers and the Quality of Care in America." Final report of the National Child Care Staffing Study. Oakland, CA: Child Care Employee Project, 1989.

Zaslow, Martha J. "Variation in Child Care Quality and Its Implications for Children." *Journal of Social Issues,* vol. 47, no. 2, 1991, pp. 125–138.

Zill, N., and J. West. *Findings from the Condition of Education 2000: Entering Kindergarten Children.* (NCES 2001-035). Washington, DC: National Center for Education Statistics, 2001.

Zill, Nicholas, and James L. Peterson. "Behavior Problems Index." Washington, DC: Child Trends, Inc., 1986.

Zimmerman, I. L., Steiner, V. G., & Pond, R. V. (1992). *Preschool Language Scale – 3*. San Antonio, TX: Psychological Corporation, 1992.

Zimmerman, I.L.S., G. Violette, and R.E. Pond. *Preschool Language Scale, Fourth Edition*. San Antonio, TX: Psychological Corporation, 2002.

This page is left blank for double-sided printing.

This page is left blank for double-sided printing.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ WASHINGTON, DC

MATHEMATICA
Policy Research