Toddlers in Early Head Start: A Portrait of 2-Year-Olds, Their Families, and the Programs Serving Them

Technical Appendices
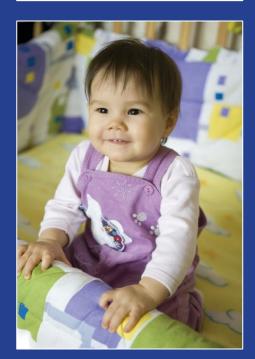
February 2015

Baby FACES 2009

OPRE Report 2015-10

Toddlers in Early Head Start:
A Portrait of 2-Year-Olds, Their Families, and the Programs Serving Them
Volume 2: Technical Appendices
OPRE Report #2015-10

February 2015

Submitted to:
Amy Madigan,
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
Jaime Thomas
Eileen Bandel
Yange Xue
Juliette Henke
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

Suggested citation: Vogel, Cheri A., Pia Caronongan, Jaime Thomas, Eileen Bandel, Yange Xue, Juliette Henke, Nikki Aikens, Kimberly Boller, and Lauren Murphy. Administration for Children and Families. *Toddlers in Early Head Start: A Portrait of 2-Year-Olds, Their Families, and the Programs Serving Them*. OPRE Report #2015-10, Washington, DC. Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. 2015.

Disclaimer: The views expressed in this publication do not necessarily reflect the views or policies of the Office of Planning, Research and Evaluation, the Administration for Children and Families, or the U.S. Department of Health and Human Services.

This report and other reports sponsored by the Office of Planning, Research, and Evaluation are available at http://www.acf.hhs.gov/programs/opre.

CONTENTS

APPENDIX A. SAMPLE AND WEIGHTS

Current Sample and Attrition

In spring 2009, we selected 1,217 children into the Baby FACES sample. From this sample, 109 children were ineligible for the study,[1] and we did not receive parental consent for 132 eligible children. Therefore, we had 976 children in the study at baseline. By spring 2010, we obtained consent for six additional children (four in the 1-year-old Cohort and two in the Newborn Cohort), but consent was rescinded for three 1-year-old Cohort children who were part of the baseline sample. In addition, two 1-year-old Cohort children previously thought to be eligible for the study were found to have ineligible birth dates; this finding, however, did not affect the sample size, because we had not received parental consent for either child. Therefore, the first spring 2010 follow-up sample has 979 children. Table A.1 presents the baseline and follow-up sample with eligibility and consent status.

Table A.1. Eligible and Consented Sample Sizes at Baseline and Follow-Up

| | | 2010 | | | |
		Eligible Consented	Eligible Nonconsented	Ineligible	Total
	Eligible Consented	973	3	0	976
	Eligible Non Non-consented	6	124	2	132
	Ineligible	0	0	109	109
2009	**Total**	**979**	**127**	**111**	**1,217**

Among the 979 eligible children in our sample with parental consent, 264 (27 percent) exited the Early Head Start program between the 2009 and 2010 data collection periods, and 715 remained. Of these 264 children, 20 had recently exited the program; that is, they had exited in either the month of or the month before our site visit. These 20 children were considered part of the sample for spring 2010, and we attempted to collect all data for them, as well as the other 715 children. The 244 who had exited the program earlier than this group of 20 were considered part of the sample for only the parent exit interview. Table A.2 presents the remaining sample and children who exited early.

Table A.2. Sample Size at Spring 2010 and Exiters

Cohort	Still in Early Head Start	Recent Exiter	Exiter (Out of Sample)	Total
Newborn	131 (67%)	4 (2%)	61 (31%)	196
1-year-old	584 (75%)	16 (2%)	183 (23%)	783
Total	**715 (73%)**	**20 (2%)**	**244 (25%)**	**979**

---

[1] Children were ineligible (either at baseline or later) if they were not actually enrolled in the program or if their birth date falls outside the specified windows. The Newborn Cohort included pregnant women whose due dates were no more than two months after the week of the spring 2009 site visit and babies whose birth dates were no more than two months before the spring 2009 site visit. The 1-year-old Cohort included children who were 10 to 15 months of age at the time of the spring 2009 site visit.

New Child-Level Weights

After the spring 2010 data collection and processing, we consulted with the Administration for Children and Families (ACF) about constructing child-level weights.[2] We decided to focus primarily on age-specific weights, and to focus each weight either on child-level data (such as staff child reports and child assessments) or staff-level data (such as staff interviews and observations). Staff can be either center-based or home visitors, and observations can be conducted either in a classroom or during a home visit. We also decided to construct preliminary age 2 weights for 1-year-old Cohort children at age 2, because 80 percent of our sample is from this older cohort and we base our main findings at age 2 on them. These weights will be finalized when data is collected on the Newborn Cohort at 2 years old in spring 2011, allowing us to make statements about all 2-year-olds regardless of cohort. Table A.3 specifies data-collection times for both cohorts at each age.

Table A.3. Ages of Sample at Data Collection Points

Cohort	Spring 2009	Spring 2010	Spring 2011	Spring 2012
Newborn	age 0	age 1	age 2	age 3
1-year-old	age 1	age 2	age 3	n.a.

n.a. = not applicable.

To construct the spring 2010 child-level weights, we first constructed a base weight, W1P, to use in the child-level weighting adjustments (see Table A.4). To construct W1P, we first ran forward and backward stepwise logistic regressions by cohort and time (1-year-old Cohort in 2009, 1-year-old Cohort in 2010, and Newborn Cohort in 2010) to predict a "complete," which indicates both parental consent and at least one parent interview. The pool of independent variables comprised the program's size stratum, service type (home, center, or mixed), urbanicity (metropolitan statistical area [MSA] versus non-MSA), and U.S. census region. We entered into a final logistic regression model variables that were common to the final models from both the forward and backward procedures. We used the inverse of the propensity score as the weighting adjustment, and applied it to the child base weight, which accounted for program selection probability; program eligibility and participation; and child (sibling) selection probability. By applying this adjustment, we weighted the consented children with a parent interview to reflect all eligible children (by age and time). To create the age 1 W1P weight, we appended the W1P weights from spring 2010 for the Newborn Cohort to those from spring 2009 for the 1-year-old Cohort.

All children with a positive value for W1P had demographic data from the parent interview that could then be used for nonresponse adjustments in the other weights. The remaining weights described in Table A.4 were constructed in a similar manner as W1P, with two major differences: (1) we used W1P as the base for these new weighting adjustments, and (2) there was a larger pool of independent variables (including sociodemographic variables from the parent interview) to use for the stepwise logistic regression models when predicting response.

---

[2] Program-level weights were constructed in 2010 the same way as baseline, accounting for each program's probability of selection, study eligibility, and participation. These weights can be used for analysis at the program level of the 89 programs.

Table A.4. Child-Level Weights

Weight Name	Age(s)	Cohort(s)	Number with Positive Weight	What Constitutes a Complete?
W1P	1	Newborn +1-year-old	839	Any parent interview
W2P_1	2	1-year-old	566	Any parent interview
W1C_1	1	1-year-old	678	Parent interview and staff-child report at age 1
W1S_1	1	1-year-old	699	Parent interview and staff interview or observation at age 1
W1C	1	Newborn +1-year-old	798	Any parent interview and an age 1 staff-child report
W1S	1	Newborn +1-year-old	825	Any parent interview and an age 1 staff interview or observation
W2C_1	2	1-year-old	562	Any parent interview and an age 2 child assessment or staff-child report
W2S_1	2	1-year-old	561	Any parent interview and an age 2 staff interview or observation
WL2_1	1+2	1-year-old	435	Parent interview at ages 1 and 2

Note: As an example, to analyze the 1-year old Cohort's vocabulary proficiency at baseline, use weight W1C_1, which is positive if the child had both a parent interview and a staff child report in spring 2009.

Table A.5 shows the sum of the child weights by cohort and data collection period, using updated eligibility information as of spring 2010. These values best estimate the number of study-eligible children being served by study-eligible Early Head Start programs.

Table A.5. Sum of Child Weights, by Cohort and Data Collection Period

Cohort	Spring 2009	Spring 2010	Age 1
Newborn	1,157	885	885
1-year-old	5,058	4,130	5,058
Total	**6,215**	**5,015**	**5,943**

We created a few additional weights for analyzing families who exited the program early and the Family Services   Tracking (FST) data. Because analysis of exiting behavior and FST data spans the year between the first two data collection points (spring 2009 through spring 2010), we required that the children have parental consent throughout this entire period. Out of 1,106 eligible children (224 in the Newborn Cohort and 882 in the 1-year-old Cohort), 973 (194 in the Newborn Cohort and 779 in the 1-year-old Cohort) met the consent criteria. Children for whom we obtained consent after baseline and those who rescinded consent after baseline were excluded from the sample and the reference population for these weights. For these weights, we did not require a parent interview; therefore, other than study cohort, the variables available to use as weighting covariates were limited to those at the program level (program size, service type, MSA status, and census region).

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 exit weight. This weight adjusted for the program's probability of selection and its participation, and whether the child had parental consent. Unlike other weights that involve spring 2010 data, we did not exclude children who left the program, as they are of key interest for this analysis. We constructed this weight separately by cohort using stepwise regression to find which of the four program-level variables predicted consent. We then used a logistic regression model to estimate a

consent propensity score; we used the inverse as a weighting adjustment. The consent-adjusted weight for the 973 consented children sums to 6,215 (1,157 for the Newborn Cohort and 5,058 for the 1-year-old Cohort).

Among the 244 children who had exited the program (but not during the month of or preceding the spring 2010 data collection visit[3]), we attempted to collect a parent exit interview. This interview aimed to collect information on where the child went after leaving the Early Head Start program, and why he or she left. We constructed a weight to adjust for nonresponse (during the first year) to this interview among those who exited. To construct this weight, we started with the consent-adjusted weight described above and excluded those who had not exited the program or had recently exited. We adjusted this weight separately by cohort. We used the stepwise logistic regression procedure described above, except that the dependent variable was whether we obtained a parent exit interview. The adjusted weight for the 128 children with a completed exit interview (33 from the Newborn Cohort and 95 from the 1-year-old Cohort) sums to 1,354 (314 for the Newborn Cohort and 1,040 for the 1-year-old Cohort).

Finally, we constructed a weight for use with the FST data. 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 year. The sum of the FST weights is 6,215 children (1,157 for the Newborn Cohort and 5,058 for the 1-year-old Cohort).

---

[3] There were 20 children who exited during the month of or preceding the site visit. We do not consider these children to have exited early for any of the weights.

APPENDIX B. DATA COLLECTION

This appendix details the process we followed to maintain relationships with study programs, manage samples, collect spring 2010 data, and prepare the data for analysis.

Baby FACES Coordinators Maintained Relationships with Study Programs and Families

After the spring 2009 data collection, Baby FACES coordinators (BFCs) continued their relationships with on-site coordinators (OSCs) during the summer and fall of 2009 with periodic telephone calls and e-mails. The purpose of these communications was generally to encourage participation in Family Services Tracking (FST), but it also reminded OSCs of the upcoming spring data collection.

In early December 2009, we mailed holiday cards to all parents of children in the study. Shortly after, we sent cards to all program directors, OSCs, teachers, and home visitors. Parents received an additional insert listing our toll-free number, encouraging them to call in and provide new contact information. The holiday cards reminded all participants of their involvement in Baby FACES and helped with locating families who may have changed addresses. The post office returned several cards with forwarding addresses that allowed us to update Mathematica's sample management system (SMS) before spring 2010 data collection.

BFCs Collected Updated Sample Information and Confirmed Site Visit Weeks

To prepare for 2010 data collection, BFCs attempted to obtain updated information on all study participants. We designed a roster confirmation spreadsheet that contained name and contact information of each study child and parent at the program, the name of each child's teacher or home visitor, each child's service type, and an exit date if the family had left the program. BFCs e-mailed these spreadsheets to OSCs in January and February of 2010 and asked OSCs to (1) review and update the address and phone number of each parent; (2) confirm each study child's service option (e.g. home- or center-based); (3) confirm each study child's teacher or home visitor; and (4) determine if and when families left the program. Further, the BFCs proposed a week for the upcoming site visit and asked the OSCs to confirm the visit. We attempted to maintain the same visit week from 2009 to 2010.[4] Once OSCs returned the spreadsheets, BFCs added to the SMS new phone numbers and addresses and updated teacher and home visitor information, if necessary; this information was used to complete parent interviews and conduct in-home child assessments.

Another important component of data collection preparation was obtaining renewed approval from institutional review boards (IRBs). Six programs had required local IRB approval during the first round of data collection. Beginning in February 2010, the BFCs assigned to these programs worked with the OSCs to obtain any necessary IRB renewals.

We Made an Additional Attempt to Obtain Missing Consent

We made one final effort to obtain consent from families who did not respond to our requests during the previous spring. In January 2010, we mailed consent packets to OSCs for distribution to

---

[4] We were able to confirm the same visit week for 74 sites and scheduled 14 visits one week before or one week after the week of the 2009 visit. One site was scheduled three weeks before the 2009 site visit week, and one site was scheduled two weeks after the 2009 site visit week.

these families, and we asked OSCs to return signed consents to Mathematica. BFCs then followed up with OSCs during the consent-gathering process to check on their progress. We sent consent packets for 55 families at 19 program sites, obtained consent for 6 additional families, and received refusals from 3 families.[5] Several families had already left their Early Head Start programs, either before or after the baseline data collection.

Training and Quality Assurance

Spring 2010 data collection included in-home child assessments and video-recorded interactions with the 1-year-old Cohort. To prepare for training and data collection, Mathematica staff pretested the child assessment procedures. In addition, we again trained and certified a set of "gold standard" service quality observers, and hired and trained field assessors.

Mathematica Staff Pretested the Child Assessment

During early fall 2009, the Baby FACES team conducted 26 pretests of the in-home direct child assessment procedures in preparation for data collection with 2- and 3-year-old children and their parents. The goal of the pretest was to assess the overall feasibility of the proposed measures, protocols, and materials in realistic and natural settings and to obtain precise estimates of the overall length of the visit and of each component. The pretest informed our assessment decisions for spring 2010 data collection with 2-year-old children and for future administrations with 3-year-olds. We video-recorded the visits to include examples of item administration and responses later in field assessor trainings and used the experiences from the pretest to prepare for that training.

We recruited pretest families with 2- or 3-year-old children through a local nonprofit organization serving low-income families. For the Spanish-speaking sample, we worked with a local nonprofit child care organization serving primarily low-income Spanish-speaking families. Any child who had exposure to Spanish (as identified by his or her parents) received the bilingual assessment; the remaining children received the English-only assessment.

The in-home assessment pretests were completed by six Mathematica staff members with child assessment experience, working in pairs. One person administered the protocol, and the second recorded the session with a video camera. The task leader (a developmental psychologist with expertise using the measures) provided 25 hours of training and practice to the Mathematica assessors on the in-home protocol and certified them prior to conducting pretest assessments. All components of the assessment were administered using paper-and-pencil administration, the same method used in the field.

During the assessment, assessors explained to the parent and child the purpose of the visit, obtained parental consent, and gave the parent a set of self-administered questionnaires to complete (including the MacArthur-Bates Communicative Development Inventories [CDI]—Infant Short Form, the Brief Infant Toddler Social Emotional Assessment [BITSEA], and the Center for Epidemiologic Studies Depression Scale [CES-D]). The assessor then set up the Preschool Language Skills (PLS-4) and administered it in English or Spanish (if the child was in a bilingual home). At the conclusion of the PLS-4, the assessor measured the child's height and weight. The assessor then set up the video camera and conducted the parent-child Two-Bag Task and Early Communication

---

[5] Three families had not been included in the original IRB application at their site and could not be pursued to obtain consent.

Indicator (ECI) task.[6] For children in bilingual homes, the assessor then administered the English PLS-4. The assessor finally asked the parent a few closing questions about whether the child's behavior in the assessment was typical for him or her and completed the Bayley Behavior Rating Scale (BRS). In sessions with 3-year-olds, the assessor administered the Peabody Picture Vocabulary Test, 4th Edition (PPVT-IV) after measuring the child's height and weight, and then transitioned to the Two-Bag Task.

Timing for the visits varied based on the age of the child and the language of the assessment. The average in-home pretest visit was 106 minutes in length. Assessments with 2-year-olds in English were the shortest, lasting 94 minutes on average, while assessments with 2-year-olds in Spanish took the most time to complete (127 minutes, on average). In-home assessments with 3-year-olds in English took, on average, 95 minutes, and assessments with 3-year-olds in Spanish were 123 minutes, on average.

The length of the pretest visits was a concern, particularly for bilingual subjects to whom both the Spanish and English PLS-4 were administered. In consultation with Nancy Castilleja (Product Manager, Speech and Language) of Pearson Assessments, the publishers of the PLS-4, we developed a protocol for dual language learners (DLLs) to eliminate the need to administer the full PLS-4 in both English and Spanish. To capture emerging language in children who are exposed to both English and Spanish, we first administer the PLS-4 in Spanish. Once the child establishes a basal and a ceiling in Spanish, the assessor transitions to administering the English PLS-4. The administrator begins with a short warm-up conversation in English, and then administers the direct English translation of all of the items that received no credit in Spanish. The assessor continues in English until reaching an English ceiling. The revised bilingual PLS-4 protocol drastically shortened the administration of the assessment.

During the pretest, we also tested two methods of obtaining height. One method used a stadiometer (a bulky, heavy instrument for measuring height). The second method used a carpenter's triangle, tape flags, and a tape measure. Because the CDC endorses the triangle procedure for measuring children at home (see http://www.cdc.gov/healthyweight/assessing/bmi/childrens _bmi/measuring_children.html) and because our pretest found no difference between measurements using the carpenter's triangle and those using the stadiometer, we selected the triangle method. It lessens the field assessor's burden by reducing the weight and size of the materials he or she must bring to each home visit.

Mathematica Survey Staff Trained Field Assessors and Observers for 10 Days

For 10 days in late January 2010, we trained staff to conduct on-site data collection. (All field staff were returning members of the team who conducted the 2009 Baby FACES baseline data collection.) Training on the in-home child assessment took place during the first 5 days, and training on classroom and home visit observations during the second 5 days. Table B.1 presents the training agenda. We trained 15 field staff (6 bilingual) to administer only the in-home child assessments; these staff attended only the first half of training. We trained an additional 19 staff members (8 bilingual) to conduct classroom and home visit observations as well as in-home child assessments; they attended all 10 days of training. Prior to the training, we mailed field staff a package of preparation materials. The mailing included the field training manual; a DVD of a home visit conducted during the child assessment pretest; and practice exercises for field staff to complete.

---

[6] Parents played with their child in the Two-Bag Task, and the assessor served as the play partner for the ECI task.

Table B.1. Training Agenda—Spring 2010

Day 1	Introductions and overview of the Baby FACES study
	ECI and Two-Bag Task introduction and practice; working with the video camera
Day 2	Height and weight training
	Overview of entire in-home visit protocol including parent SAQ and observational items
	PLS-4 training
Day 3	PLS-4 training and practice
Day 4	Practice conducting child assessment with toddlers
	Bilingual child assessment training
Day 5	Child assessment certification
Day 6	CLASS-T training
Day 7	CLASS-T reliability
Day 8	CLASS-T in-field reliability observation
	ITERS-R review
Day 9	ITERS-R in-field reliability observation
	HOVRS-A review and recertification
Day 10	CLASS-T practice
	ITERS-R reliability

In-home child assessment training. The topics covered during in-home child assessment training included scheduling appointments; setting up and properly using the video camera; measuring height and weight; administering the PLS-4 assessment and ECI and Two-Bag interactions; and a general session called "Working with 2-Year-Olds." On the fourth day of training, field staff practiced the entire child assessment, including an abbreviated administration of the PLS-4. On Day 5 of training, the staff conducted full child assessments including the full PLS-4, attempting certification in the following areas: (1) taking height and weight measurements; (2) setting up and successfully recording the Two-Bag interaction; (3) administering and successfully recording the ECI; and (4) conducting the PLS-4.

To measure adherence to the step-by-step protocol instructions, Mathematica staff members who were trained to serve as gold standard assessors for child assessment certification, observed each field staff member during both administrations of the child assessments, with a certification form that we developed. Following the second administration, certifiers provided trainees with detailed written feedback on their performance. Most staff earned provisional certification at the training and full certification upon Mathematica review of a video-recorded assessment of a 2-year-old post training.[7]

Gold standard assessors reviewed the post-training child assessment certification videos for 31 trainees and provided trainees with detailed feedback.[8] Based on their first videos, 25 field staff

---

[7] Each field staff member was required to conduct and record the entire assessment with a 2-year-old and send in the recording for review and final certification before conducting assessments in the field.

[8] Two of the staff trained to administer the in-home child assessment received waivers because of their exceptional performance at the training and were not required to complete a certification video.

became certified; 5 were certified on a second attempt. One field staff member was not certified but remained on the project as a classroom and home visit observer (for this area, he had obtained certification). The final count of field staff included 18 observers/assessors (8 were bilingual), 1 observer only, and 12 staff serving as only assessors (4 were bilingual). (Two staff members left the study before data collection began.)

Classroom and home visit observation training. We trained the 19 field staff observers to conduct the Classroom Assessment Scoring System-Toddler (CLASS-T), Infant Toddler Environmental Rating Scale-Revised (ITERS-R), and Home Visit Rating Scale-Adapted (HOVRS-A). On Days 6 and 7, the authors of the CLASS-T came in to train and certify all field staff. The CLASS-T author Karen LaParo trained and certified one Mathematica survey staff member before the training to serve as the gold standard observer. Observers rated video recordings during the two days of training, and we calculated inter-rater reliability. Only 6 of the 19 observers were unable to meet the 80 percent inter-rater reliability threshold. On Day 8 of training, observers conducted CLASS-T observations in classrooms, and all observers were reliable with the gold standard observations.

Because all observers were previously trained and certified in administering the ITERS-R, we reviewed the administration during field training. Before training, nine ITERS-R gold standard observers participated in at least one recertification training with Mathematica's lead trainer. Then, on Day 9 of the field training, these gold-standard observers led in-field reliability observations, resulting in recertification of all but one of the observers (for a total of 18).

These staff were also re-certified on the HOVRS-A. Approximately six weeks before the training, we mailed field staff a packet containing the HOVRS-A manual, a DVD with eight practice clips, another DVD with five clips for reliability assessment, multiple HOVRS-A scoring booklets, and a memo with instructions (and answers to the practice exercises). We instructed observers to spend approximately 10 hours reviewing the HOVRS-A manual, watching and scoring the practice videos, checking their work against the answers provided, and then watching and scoring the five reliability videos. They mailed their completed reliability observation booklets to Mathematica prior to training. Mathematica gold standard observers reviewed the booklets and prepared feedback to give to the observers at training. A group review and practice session was held at the field training and was followed by additional sessions in which observers continued to view and code video clips to achieve certification. By the end of training, all 19 observers were certified using the 80 percent reliability threshold.

Field teams. Before staff conducted any field visits, we created field teams. Field teams consisted of a team leader and one or more field observers and assessors. The team leader was responsible for managing on-site activities, including scheduling classroom and home visit observations, child assessments, and in-person interviews with teachers and home visitors. Team leaders were also the main point of contact with the OSC during the site visit week. If study children at a site were exposed to Spanish in their households, we sent at least one bilingual team member to the program site. Teams and team leaders were reconfigured each week of data collection to accommodate the needs and sizes of the programs visited.

Gold Standard Observers and Assessors Conducted Quality Assurance Field Visits

For quality assurance (QA) purposes, we sent trained gold standard observers to monitor each member of the field staff conducting the HOVRS-A, ITERS-R, and CLASS-T. Six Mathematica

staff members and two employees of Branch Associates served as the gold standard observers on 24 sites visits over nine weeks. For the HOVRS-A, the gold standard observer accompanied each field observer to monitor a home visitor conducting a home visit. The gold standard and field observers rated the visit independently and discussed their scores immediately afterward to arrive at a consensus score for data entry. Inter-rater reliability was calculated as having at least 80 percent exact agreement with the consensus score. For the ITERS-R and the CLASS-T QA, the gold standard observer and up to two observers completed the classroom observations simultaneously, and then reached a consensus score. Reliability was calculated, with a standard of 80 percent or more of the items within one point of the consensus scores. [9] For the CLASS-T, observers were considered reliable if they were within one point of the gold standard observer at least 80 percent of the time. There was no consensus scoring for the CLASS-T, but the gold standard observer and the field observers discussed each item immediately following the observations.

We also conducted a QA review of the PLS-4 portion of the in-home visit. Beginning in the fourth week of data collection and continuing through the eighth week, each assessor recorded a full administration of the PLS-4 during a regularly scheduled in-home child assessment and sent it to Mathematica for review by gold standard assessors. Feedback for those who did not pass the PLS-4 QA was sent to both the assessor and his or her team leader. The team leader then accompanied the assessor on his or her next child assessment visit to monitor progress and provide immediate feedback.

Two Teams of Video Coders Were Trained to Reliability

We created two teams of coders, each supervised by a gold standard coder; the entire coding team was supervised by a senior survey researcher. One team coded the Two-Bag Task using the Parent-Child Interaction Rating Scales for the Two-Bag Assessment. The other coded the Two-Bag Task using the Parenting Interactions with Children: Checklist of Observations Linked to Outcomes (PICCOLO), and the ECI task. Each team received extensive training and ongoing reliability checks as described below.

Two-Bag Task. A certified trainer with extensive experience coding the Two-Bag Task using the Parent-Child Interaction Rating Scales for the Two-Bag Assessment trained 13 members of the Mathematica coding team. An additional bilingual coder was later trained by the coding team leaders. Coders independently coded a video-recorded interaction coded a priori by the trainer to serve as the certification video. The certification criterion required that coders achieve 92 percent agreement (exact or within one point) with the ratings assigned by the trainer across the 12 scales. Two additional certification videos were available to coders who did not certify on the initial certification video.

Of the original team of coders who were trained by the certified trainer, six members composed the final coding team. Following training and certification, two trained team leaders worked with the six-member coding team to establish and maintain inter-rater reliability throughout the coding period. Inter-rater reliabilities between the team leaders and coding team members were established

---

[9] Although infrequent, a gold standard score can differ from the consensus score. This situation would occur if, during the debriefing session, the observer produces sufficient evidence to cause the gold standard observer to reassess the original score (for example by observing something that the gold standard observer did not see). This revised score becomes the consensus score. The observer's reliability is then calculated based on the consensus score, because that data is the most accurate.

on the 12 seven-point scales to a criterion of 80 percent, allowing for a one-point difference in scores. Thereafter, the team conducted weekly inter-rater reliability checks on three to five randomly selected videos.

PICCOLO. For the Baby FACES study, a certified trainer with extensive experience coding the Two-Bag Task using the PICCOLO trained 13 members of the Mathematica coding team. Coders independently coded four video interactions coded a priori by the PICCOLO research team to serve as the certification videos. Coders were required to achieve 80 percent exact agreement with the trainer's scores across the four videos. Notably, attaining 80 percent exact agreement with the developers' ratings proved to be a challenge for the coding team. Exploratory analysis of the team members' ratings indicated that the most common discrepancy occurred when distinguishing between scores of "0" (behavior was not observed) and "1" (behavior was infrequently observed) on the three-point scale. In close consultation with and at the recommendation of the measure developers (Roggman, personal communication, April 2, 2010), we pursued certification using a binary scale in which "0" and "1" were collapsed to represent behaviors that were absent or infrequently observed. A score of 2 still indicated behaviors that were "clearly evident" and frequent in their occurrence and/or intensity. According to this revised criterion, team leaders achieved 80 percent exact agreement with the trainer's scores across the four certification videos. To facilitate certification efforts and to aid in establishing and maintaining inter-rater reliability among members of the coding team, certified team leaders provided additional detail and examples to clarify the coding definitions and anchor the scale points for each of the PICCOLO's 29 items.

Early Communication Indicator. Prior to the Baby FACES spring 2010 data collection, two expert consultants trained 13 members of the Mathematica coding team on the ECI. Coders were required to become certified on two videos coded a priori by the developer. The certification criterion required that coders achieve 85 percent agreement with the developer ratings on each video. Coders recoded the videos as many times as necessary until meeting the certification criterion. Of the original team of coders who were trained by the ECI expert consultants, eight members composed the final coding team. Following training and certification, two trained team leaders worked with the eight-member coding team to establish and maintain inter-rater reliability throughout the coding period.

Telephone Interviewers Were Again Trained to Administer the Parent Survey

In 2010, we followed the same training process for the parent survey as in the 2009 Baby FACES data collection. Two groups of (daytime and evening) telephone interviewers received eight hours of training for the parent survey in early February 2010. We trained 16 telephone interviewers (four were bilingual in English and Spanish). In addition, four monitors (two of whom were bilingual) and three telephone supervisors participated in the training sessions. Training involved a brief overview of the project and how the parent interview fit into the overall data collection effort, instruction on gaining cooperation and screening of parents, and a question-by-question review of the survey instrument. At the conclusion of the formal training, interviewers were paired up to conduct mock interviews with one another using the Computer-Assisted Telephone Interview (CATI) instrument under the guidance of the trainer and supervisors. During the first weeks of telephone interviewing, each interviewer was monitored and given immediate feedback. Ongoing monitoring of 10 percent of the interviews continued throughout the telephone field period. We monitored bilingual interviewers in both English and Spanish.

Training on Program Director Survey

In late March 2010, we reassembled the team of researchers that conducted the program director interview in 2009. One additional researcher was added to the team for 2010. The Baby FACES project and survey directors met with the two returning and one new researcher for a four-hour training session. The training built on what the researchers had learned from the 2009 survey, highlighting areas that sparked questions the year before and describing the intent of new questions added for this round. In addition, the training included a review of the programs' structure and stressed the importance of gathering enough information from the program director to accurately record the information on the questionnaire. Extensive spreadsheets were created to capture additional information provided during the semi-structured interviews about programs' organization and activities.

Interviews, Observations, and Assessments

Mathematica Conducted Exit Interviews with Families Leaving Early Head Start

Between the end of baseline data collection in 2009 and the beginning of follow-up data collection in 2010, approximately 20 percent of the study children stopped attending or receiving services from their Early Head Start programs. By the time data collection began, the rate was close to 25 percent. This rate of attrition was higher than expected, and we wanted to better understand why families leave Early Head Start and where they go once they do leave.

The Baby FACES research team periodically conducted exit interviews. We conducted the first round in fall 2009 and a second round at the same time as the parent interview during spring 2010. The 20-minute telephone interview was designed to collect (1) the child's exit date from the perspective of the parent; (2) the reasons for leaving the program; (3) satisfaction with the program; (4) child health; (5) current child care arrangements; and (6) annual household income and sources of income support. We offered parents $20 to complete the interview.

Exit interview, round one. In late October 2009, six interviewers, two monitors, and one supervisor received exit interview training for four hours. The training provided a brief overview of the interview, a question-by-question review of the survey, and practice in pairs. The exit interview was designed to be conducted by telephone and coded on paper. Therefore, much of the training was devoted to following basic skip patterns on the paper instrument.

As part of exit interview preparation, BFCs asked OSCs to provide rosters listing all children currently enrolled in their programs, with enrollment dates and dates of birth. The rosters helped us to confirm any child who had exited or returned to the Early Head Start program since the spring 2009 data collection. We began receiving rosters in September 2009 via e-mail, fax, and posted mail. Because roster formats differed across sites, one BFC reviewed all rosters for quality assurance to confirm dates for enrollment and exits and to resolve discrepancies. By December 2009, we had received rosters from 83 programs.

Because the rosters took some time to receive, we started conducting the exit interview in October 2010 with families that we knew had left the program. A second release of sample was added in November, when additional families were identified as leavers based on the rosters received to that point. Because we planned to conduct a second round of exit interviews during spring 2010 data collection, we attempted exit interviews for cases where we had received a fall roster before or

during the first two weeks of data collection.[10] The first wave (October) contained 91 cases from 51 sites; the second wave (November) contained 25 cases from 10 sites. The first round of exit interview data collection ended in December 2009.

Two families turned out to be ineligible for the exit interview because the parents informed interviewers that their children were still enrolled in the Early Head Start study programs. BFCs confirmed with the OSC that the children were still enrolled in the program, and these cases were then marked as eligible for spring 2010 data collection.

Exit interview, round two. The second round of exit interviews began in March 2010 at the same time as the parent interview. It contained exit cases not collected during the first round because we had not received a fall roster. It also contained new exits we learned about from (1) roster spreadsheets being returned to BFCs prior to spring data collection; (2) staff entering exit dates into the FST; (3) parents informing interviewers during the parent interview; and (4) field staff notification during site visits. In addition, we decided to make a second attempt to interview parents who did not complete exit interviews during round one.

Data collection began mid-March 2010 and finished at the end of June 2010. The total sample to be interviewed in that period was 203 parents, including 52 eligible cases that did not complete the first round of the exit interview the previous fall.

During round two of exit interviews, seven cases turned out to be ineligible for the exit interview because parents informed interviewers that their children were still enrolled in their Early Head Start programs. In some instances, the exit information provided by the programs was inaccurate or misinterpreted. In other instances, the child left briefly, then returned to the study program; we decided to consider those children in-sample.

A large portion of the exit sample was hard to reach. When we could not reach parents because all phone numbers on record were not working, we attempted online search engines such as Reach 411 and Accurint. However, these searches provided minimal new contact information. During the baseline parent interview, though, we had asked parents to provide the names, phone numbers, and addresses of three people who would know how to reach them in the future. Many of these contacts were indeed reachable and offered updated location information for several parents.

Because the attrition rate was higher than we expected, we decided to include in the 2010 data collection families who recently stopped receiving Early Head Start services if they exited near the time of their site visit weeks. We based our definition of "near the time of the visit" on the fact that we received exit dates from multiple sources (roster confirmation spreadsheets, OSCs, field staff, and parent interviews), making it difficult to assign an exact "exit" date. Parents and programs were often able to provide only an exit month. Therefore, our eligible data collection window included cases with exit dates in the month of the site visit or the month before the site visit. For example, if the first day of the site visit was scheduled for March 15, then any exit date from February 1 to March 14 was considered inside the data collection window. Any exit date before February would be considered outside the data collection; these parents would instead be contacted to complete the exit interview.

---

[10] Of the 83 rosters received, 67 were in time for data collection.

Recently exiting families considered eligible for regular data collection were treated the same as families still receiving Early Head Start services from the study program. We attempted the in-home child assessment if the family was still in the program area and still called to complete the regular parent interview. We also asked teachers and home visitors to fill out a staff-child report (SCR) and complete the teacher/home visitor interview. If a child was in center-based care before leaving the program, we conducted an ITERS-R or CLASS-T observation of the child's last classroom. If a child was in home-based care and he or she was the only study child on the home visitor's caseload, we were unable to complete a HOVRS-A with the home visitor. However, we attempted observations if the home visitor served other study families.

On-Site Data Collection Consisted of Classroom and Home Visitor Observations, Teacher/Home Visitor Interviews, and In-Home Child Assessments

Teams of field assessors visited 89 sites over 16 weeks from March 2010 through mid-June 2010.[11] Several weeks before the site visits, BFCs and Survey Operations Center (SOC) field supervisors developed data collection plans (DCPs) for each site. The DCPs serve as the team leader's guide to collecting all required instruments. Each DCP included (1) the addresses and contact information of program centers and OSCs; (2) a listing of each study child including his/her teacher or home visitor, service type, cohort, and exposure to Spanish at home; and (3) a listing of teachers/home visitors, including a checklist of the required instruments.

Classroom observations. We conducted observations of all teachers using the ITERS-R and the HOVRS-A. The ITERS-R observation was again used to observe all teachers providing services to Newborn Cohort children, including recording information on space and furnishings, personal care routines, listening and talking, activities, and program structure. In 2010, we also used the CLASS-T to observe all teachers providing services to children in the 1-year-old Cohort. The CLASS-T collects information on classroom climate (both positive and negative), teacher sensitivity, regard for child perspectives, behavior guidance, facilitation of learning and development, quality of feedback, and language modeling.

Ten teachers serving both a Newborn and a 1-year-old Cohort study child were observed using both the ITERS-R and CLASS-T. Only in these instances was a teacher observed more than once. If only one observer was on site, he or she conducted the two observations on different days. Otherwise, two observers conducted the ITERS-R and CLASS-T at the same time.

We conducted classroom visits similarly to the 2009 data collection. Whenever possible, we scheduled the classroom observations in the morning, when children are most active. Only a few observations were conducted in the early afternoon. During each observation (ITERS-R and CLASS-T), the observers completed two counts of children and adults (spaced at least an hour apart) and completed three post visit ratings; there was no interaction with the children or the teachers. At the end of the observation, the observers asked the teachers about events that were not observed (for example, "Since I was not here at naptime, can you please describe how nap is handled?"). After gathering these last pieces of information, the observers assigned their final scores. Classroom observations generally lasted two to three hours. We gave a gift bag of classroom supplies worth $25 to the teacher in each observed classroom.

---

[11] Assessors visited between 1 and 10 sites during each week of data collection (there were two weeks with no site visits scheduled).

Home visitor observations. Using the HOVRS-A, observers conducted a home visit observation of each home visitor serving at least one child in the study, the same as in the 2009 data collection. This observation tool focuses on the quality and nature of aspects of the home visit interaction, including home visitor responsiveness to the family, the relationship between the home visitor and the parent, and the engagement of the parent and the child during the home visit. During these observations, observers did not interfere with or react to home visit activities or interactions. Each HOVRS-A observation lasted an average of 90 minutes. We gave each observed home visitor a gift bag of classroom supplies that was identical to those we gave the teachers.

Field assessors were required to conduct in-home child assessments with children in the 1-year-old Cohort. Therefore, we attempted to conduct the home visit observation with families of children in the Newborn Cohort, who would not be receiving the in-home child assessments. When this option was not possible, families were by and large amenable to receiving both the observation and in-home child assessment. Nearly all families of children in the 1-year-old Cohort receiving a HOVRS-A observation also completed the in-home child assessments. About half of parents requested that team members conduct both the child assessment and the home visit observation on the same day. When a home visit could not be conducted during the field visit week, because the family receives visits every other week, field supervisors sent team members back the next week to complete the HOVRS-A.

Teacher/home visitor interview and staff-child report. We again interviewed each teacher and home visitor providing services to at least one study child. The in-person interview was nearly identical to the spring 2009 interview. It lasted approximately 30 minutes and focused on their background, training, services they provide to families, and expectations the programs place on them. Minor changes to the interview included two new questions about smoking and some modifications to age-specific questions (since the children were now 1 and 2 years old).

The field assessor conducted the interview and recorded the teacher/home visitor's responses on a paper questionnaire for later data entry. The two instruments were nearly identical, with a few questions modified to reflect services provided in the home versus in a classroom.

At the start of the week of the site visit, we again distributed SCR forms to each teacher and home visitor of study children and instructed them to complete this self-administered questionnaire (SAQ) about each of the study children. In 2010, there were four versions of the SCR: one each for home visitors and teachers of children in the Newborn and 1-year old Cohorts. (That is, there was a home visitor form for Newborn Cohort children and another for 1-year-old Cohort children, and so on.) Teachers were asked to report on two additional questions specific to child language exposure in the classroom. All teachers and home visitors were asked to report on the child's social skills, language development, and parent-staff relationships. The key difference between the Newborn and 1-year old Cohorts' SCRs was the version of the language measure (CDI) used; teachers and home visitors of the Newborn Cohort were asked to report on a different set of words than those of the 1-year old Cohort.

After the visit, the team leader collected and reviewed all completed instruments and sent the documents to Mathematica for receipt and review. Whenever possible, we collected completed SCRs before the end of the visit week. The forms took about 15 minutes to complete for each child. Teachers and home visitors received $5 for each completed form. If the SCRs were not delivered or completed by the end of the site visit, teachers and home visitors were given prepaid business reply envelopes and were asked to mail the documents to the SOC.

Child assessments. During spring 2010, assessors conducted in-home child assessments with each 2-year-old study child. The assessors attempted to schedule the assessments with parents a week in advance of the site visit, when possible. Approximately two weeks prior to the target week, we sent parents a letter explaining the visit and included a list of activities the parent could "try out" with the child to prepare for the visit (see Parent SAQ below). When parents completed the telephone interview in advance of the site visit, assessors provided them with a toll-free field number they could call to schedule the assessment and obtain more information. If an assessment was not scheduled ahead of the site visit, the team leader worked with the OSC and teacher or home visitor to contact the family and find an appropriate date. Occasionally, depending on the comfort level and/or availability of the parent, the program provided a private setting at a center to conduct the assessment.

The 24-month Child Assessment Record Form provided the assessor a script to follow and coding pages to record responses and observations. There were two versions of the Child Assessment Record Form: an English version and a bilingual version (containing both English and Spanish items). If our records or previously collected data indicated the child is exposed to Spanish in the home, we sent the assessor a bilingual record form. If we did not have this information, the assessor asked the parent questions (during the scheduling call or at the beginning of the visit) to determine which form to use.

The first activity performed with the child was the PLS-4 scale, which is used to determine the child's language comprehension skills. Children are asked to follow directions and perform tasks using manipulatives (such as a box, toys, and blocks) or point to objects found in a picture manual. Assessors had the 24 Month Baby FACES PLS-4 Flip Book, which provides step-by-step instructions for administering each item of the PLS-4. We created two versions of the Flip Book: one for the English PLS-4 and one for the Bilingual PLS-4. The PLS-4 took approximately 20 minutes to conduct (slightly longer for bilingual children who are asked some questions in both languages).

The PLS-4 was followed by the collection of the child's height and weight. The assessor twice measured the child's height in centimeters and weight in kilograms. If the two measurements differed by more than 0.2 kg or 2 cm, assessors took a third measurement. When measuring height, the assessor asked the child to stand straight up (without interference) against a wall or door. The assessor then lowered a carpenter's triangle from above the child to where it firmly touched the crown of the child's head. The assessor placed a removable stick-on arrow flag on the wall at the bottom of the triangle to mark the child's height, then used a metal measuring tape to determine the child's height. To record the child's weight, the assessor used a digital scale.

Next, the assessor recorded the parent-child Two-Bag Task. The purpose of the interaction is to assess parent and child behaviors as the pair interact in semi-structured free play. To set up the interaction, the assessor first found an unobtrusive place for recording, placed a yoga mat on the floor, and set up the video camera. Then, the assessor began recording by holding a signboard in front of the camera for 15 seconds (to identify the video). Once set-up was complete, the assessor asked the parent and child to sit on the mat. The assessor then placed bag number 2 on the floor to the parent's left and bag number 1 on top of it. The first bag contained a book entitled *Goodnight Gorilla*, and the second bag contained a set of toy dishes and play food. The assessor, using scripted standardized instructions, informed the parent that the recording would last eight minutes and that her or she may play with the child. The assessor asked the parent to begin playing with bag number

1 first, then move on to bag number 2 whenever he or she liked. During the recording, the assessor monitored the video camera to verify that the parent and child stayed within the video frame.

After completing the parent-child interaction task, the assessor set up the observer-child task, the ECI. The purpose of this task is to measure the child's expressive communication skills. With the video camera running, the assessor changed places with the parent on the mat, unpacked a Fisher-Price farm set, and interacted with the child for six minutes. Per the ECI developer training standards, the assessors were trained to follow the child's lead, talk about things of interest to the child, comment on the child's actions, and repeat what the child was saying (particularly if the child spoke softly and the microphone might not pick up their vocalizations). The assessor was instructed to ask questions sparingly. Following the video tasks, the assessor gave the child a *Goodnight Gorilla* book as a gift.

The assessor finished the visit by asking the parent a series of questions intended to help researchers understand the context of the visit. The questions ask whether the child's behavior during the visit was typical, whether the child is generally shy or outgoing, and whether the parent thought the child did his or her best. The assessor then collected the self-administered questionnaire that the parent had received at the start of the visit (see Parent SAQ below). Next, the assessor verified that the parent had completed the telephone interview. If the interview had not been completed, the assessor offered the parent his or her cell phone and called the SOC to complete the interview. If the parent did not wish to complete the interview at that time, the assessor gave the parent our toll-free number and called to set up an appointment with the SOC. The assessor then provided the parent with an appointment card that included the toll-free number and time and date of the appointment. Finally, before leaving the house, the assessor gave the parent a $35 check as a thank you for both completing the telephone interview and their involvement in the child assessment. The visit lasted one and a half hours, on average.

As soon as possible after leaving the home, the assessor completed ratings of the (1) child's behavior during the session (Bayley BRS); (2) their observation of the home (using the Home Observation for Measurement of the Environment, or HOME); and (3) the neighborhood, using items from a study of neighborhoods in Chicago (Ross et al. 2008). The HOME observation measures the quality of stimulation and support available to children in their home environments, and the neighborhood items describe the condition and safety of the neighborhood.

Parent SAQ. Parents of the 1-year old Cohort study children were asked to complete a self-administered questionnaire (SAQ) during the child assessment. The SAQ consisted of several measures about child development in a number of areas, including social skills, communication, language development, gross and fine motor development, and personal growth. Some of these measures were administered at baseline during the telephone interview. As previously mentioned, parents received an advance letter that included a list of activities the parent could try out with the child to prepare for the visit. The purpose of this list was to prepare parents to complete the Ages and Stages, Third Edition (ASQ-3) portion of the questionnaire. This measure asks parents whether the child can regularly, sometimes, or not yet do particular activities. Because the PLS-4 and recorded interactions require a substantial amount of time, the intent of the advance letter was to provide parents an opportunity to try some of these activities (which parents may have never tried or seen their children do) before the visit. Some activities from the advance letter included:

- Let your child drink from a cup and use a spoon

- Let your child kick a soccer-sized ball; watch him or her run, jump, and go up and down stairs

- Let your child practice stacking with a least eight small blocks or containers

- Give your child beads or Cheerios and string to practice beading

- Read books to your child, ask questions about the pictures, and let him or her try to turn the pages

- Let your child scribble and try to make lines with crayons or markers

After handing the parent the SAQ, the assessor explained that the parent should not expect to see the child do everything listed. If the parent wanted to try out an activity, he or she was told to circle it. The parent could then come back to the activity at the end of the visit and try it with the child.

Mathematica Conducted Parent Interviews and Program Director Interviews by Telephone

Parent interviews. Parents of each study child were again asked to complete a telephone interview. To reduce the burden on parents, we shortened the interview in 2010 and moved several measures to the parent SAQ. Further, many questions (such as race and ethnicity) were asked only if we did not conduct a baseline interview and thus did not have the information on record.

Mathematica telephone interviewers conducted the interview at the SOC in Princeton, New Jersey. The interview was programmed and administered using CATI, thereby allowing the individual path of each interview to be determined based on the responses given to previous questions or preloaded information. The interview was conducted in Spanish when necessary. The parent interview had 20 sections, although not all parents were asked questions in every section. The parent interview instrument, along with all other study instruments, will be available on the Administration for Children and Families (ACF) website.

We conducted parent interviews from mid-February through June 2010. In this round of data collection, we attempted to conduct the parent interviews before the field site visits. We began calling parents at least one month prior to their site visits, when possible. Parents were sent an advance postcard informing them of the upcoming interview and reminding them of the upcoming Baby FACES data collection. We released four large groups of sample to the CATI system between February and April. We temporarily stopped calling parents during their site visit week to allow field staff to schedule appointments for the in-home child assessments.

In the screener portion of the interview, parents were asked to confirm whether they were still receiving services from the study program. If the family had left the program, the parent was asked for an exit date. The CATI program was designed to calculate whether the exit date was inside the eligible data collection window for families who left early to receive the parent interview. If it was, the interviewer proceeded to complete the full parent interview. If it wasn't, the interviewer would find an exit interviewer to complete the exit interview or set an appointment for one.

While calling parents, interviewers identified many incorrect or nonworking telephone numbers. We generally put these cases on hold and waited for field staff to locate the parents for the in-home child assessment. BFCs also played a role, by contacting OSCs and asking them for updated phone numbers. In some instances, the programs provided private space at a center where parents could

call in and complete the interview using the center phone. As a final attempt, SOC locators attempted to find telephone numbers through directory assistance and online sources.

The average length of the parent interview was approximately 55 minutes, with 1-year-old Cohort interviews running 50 minutes, on average, and Newborn Cohort interviews lasting 60 minutes. Spanish interviews lasted 60 minutes, on average. Parents received $35 for completing this round of data collection.[12]

Program director interviews. To learn about program practices, policies, and overall enrollment, we conducted interviews with program directors. The program director interview was broad in scope and asked directors about the entire program, including (if applicable) all of their Early Head Start centers (not just those selected for the study). We gathered program-level information in two ways: (1) an hour-long telephone interview, and (2) a self-administered questionnaire. While the process for conducting the interviews was similar to 2009 data collection, the 2010 versions of the instruments were revised and included new and changed questions. The interview focused on program structure, involvement with community partners, approaches to serving DLLs, implementation, and program goals. The SAQ included questions that might require review of records or consultation with others—for example, to describe the hours of operation in each of their centers, quantify the types of families enrolled in their program, and indicate the different services provided and by whom.

We conducted interviews and collected SAQs from April 2010 to July 2010. Each program director was mailed an advance letter and SAQ. We sent multiple e-mail reminders to programs that did not return the SAQ within a month of the mailing. About a week after the mailing, researchers began calling to schedule the telephone interviews at the program director's convenience.

Upon completion, each researcher reviewed his or her own interviews, entered verbatim comments into a spreadsheet, reviewed the SAQ for completeness, and determined whether a call-back was needed. Because the program director interview resembled a semi-structured executive interview, the interviewers recorded extensive additional information on spreadsheets to capture data that went beyond the questionnaire form and could facilitate better understanding of program activities.

Response Rates

Most Parents Completed the Shortened Parent Interview

We completed telephone interviews with 567 parents (77 percent). See Table B.2 for response rates. A few parents partially completed the interviews: 8 parents (1 percent) completed the household composition and household languages sections, and another 8 parents (1 percent) completed at least half of the interview. We completed 32 first-time interviews with parents we did not reach at baseline. In addition, 20 interviews were conducted with new parents, meaning someone other than the baseline respondent completed the 2010 parent interview. In most of these cases, one of the child's birth parents completed the baseline interview, and the other birth parent completed the follow-up.

---

[12] In the case of the Newborn Cohort, we mailed parents a check within two weeks of completing the parent interview. For the 1-year-old Cohort, we either gave them a check at the conclusion of the in-home child assessment or mailed a check at the conclusion of the parent interview if they did not complete the in-home assessment.

A high percentage of parents from both cohorts completed the interview: a total of 80 percent of Newborn Cohort parents and 79 percent of 1-year-old Cohort parents. We completed 121 interviews in Spanish (101 from the Newborn Cohort and 20 from the 1-year-old Cohort).

We were unable to complete the interview with 152 parents (21 percent). Of parents not completing the interview, most (53 percent) were parents with working telephone numbers for whom we continuously reached voicemail systems, failed to receive an answer, or could not set up a suitable appointment time. We were unable to locate a working telephone number for 31 percent of this group, and 16 percent refused to complete the interview. In four instances, the interview could not be completed because of a language barrier.

Assessors Achieved a High Response Rate for the Child Assessment

We conducted a large percentage of in-home child assessments and video-recorded interactions with 1-year-old Cohort children and their parents. Assessors visited the homes of 547 children (91 percent) and completed the Child Assessment Record Form. We received 145 bilingual versions of the Child Assessment Record Form. A small number of parents (3 percent) refused the entire in-home visit. We were unable to locate three families. The remaining parents were typically unable to overcome scheduling conflicts.

From the SAQs we distributed during the in-home assessment, we collected or received 537 (out of the 547 visits and 600 total cases) for a completion rate of 90 percent. In addition, we received 522 (87 percent) working Two-Bag videos and 519 (87 percent) working ECI videos. See Table B.2 for completion rates.

Teachers and Home Visitors Continued to Participate in High Numbers

Response rates remained high for teacher and home visitor instruments (see Table B.3). We observed 220 classrooms of teachers with 1-year-old Cohort children, achieving a response rate of 99 percent on the CLASS-T. Observers also completed 53 observations of classrooms with teachers of Newborn Cohort children, achieving a response rate of 98 percent on the ITERS-R. We observed 193 home visits for a response rate of 83 percent on the HOVRS-A. In addition, we completed 264 teacher interviews and 225 home visitor interviews, for response rates of 99 percent and 97 percent, respectively. Teachers and home visitors were again receptive to completing the SCRs. We received a total of 703 SCRs from 374 teachers and 329 home visitors for a 96 percent completion rate.

Table B.2. Baby FACES Consent and Child Instrument Completion Rates

Week	Date	No. of Consented Children	No. of Completed Parent Interviews	Completed Parent Interviews (%)	No. of SCRs Received	SCRs Received (%)	No. of Child Record Forms Received	Child Record Forms Received (%)	No. of Parent SAQs Received	Parent SAQs Received (%)	No. of ECI Videos Received	ECI Videos Received (%)	No. of Two-Bag Videos Received	Two-Bag Videos Received (%)
			Both Cohorts (n = 735)				1-year-old Cohort (at Age 2) Only (n = 600)							
1	3/1	41	31	75.6	40	97.6	28	93.3	28	93.3	28	93.3	28	93.3
2	3/8	71	65	91.5	69	97.2	54	90.0	54	90.0	52	86.7	52	86.7
3	3/15	82	60	73.2	81	98.8	65	92.9	60	85.7	59	84.3	60	85.7
4	3/22	91	79	86.8	90	98.9	67	90.5	67	90.5	60	81.1	60	81.1
5	3/29	22	16	72.7	22	100.0	19	95.0	17	85.0	16	80.0	15	75.0
6	4/5	25	24	96.0	24	96.0	18	94.7	18	94.7	17	89.5	18	94.7
7	4/12	58	46	79.3	57	98.3	42	85.7	42	85.7	41	83.7	40	81.6
8	4/19	26	20	76.9	25	96.2	21	95.5	21	95.5	19	86.4	19	86.4
9	4/26	78	58	74.4	73	93.6	55	90.2	54	88.5	53	86.9	54	88.5
10	5/3	75	59	78.7	68	90.7	57	95.0	57	95.0	57	95.0	57	95.0
11	5/10	56	41	73.2	56	100.0	42	93.3	42	93.3	39	86.7	41	91.1
12	5/17	74	56	75.7	67	90.5	55	88.7	54	87.1	55	88.7	55	88.7
14	5/31	21	46	76.2	18	85.7	15	83.3	15	83.3	14	77.8	14	77.8
16	6/14	15	12	80.0	13	86.7	9	90.0	8	80.0	9	90.0	9	90.0
Total		**735**	**583**	**79.3**	**703**	**95.6**	**547**	**91.2**	**537**	**89.5**	**519**	**86.5**	**522**	**87.0**

Source: Spring 2010 sample management system.

Note: Data collection did not occur during weeks 13 and 15.

Table B.3. Baby FACES Teacher Instrument Completion Rates

Week	Date	No. of Consented Children	No. of HOVRS-A Received	HOVRS Received (%)	No. of CLASS-T Received	CLASS-T Received (%)	No. of ITERS-R Received	ITERS-R Received (%)	No. of Teacher Interviews Received	Teacher Interviews Received (%)	No. of Home Visitor Interviews Received	Home Visitor Interviews Received (%)
1	3/1	41	13	81.3	13	100.0	3	100.0	15	100.0	15	93.8
2	3/8	71	28	84.8	15	100.0	3	100.0	18	100.0	33	100.0
3	3/15	82	10	62.5	41	100.0	5	100.0	45	100.0	16	100.0
4	3/22	91	29	87.9	22	100.0	7	100.0	29	100.0	33	100.0
5	3/29	22	9	90.0	5	100.0	--	--	5	100.0	10	100.0
6	4/5	25	88	72.7	6	100.0	2	100.0	7	100.0	11	100.0
7	4/12	58	14	87.5	20	100.0	5	100.0	22	100.0	16	100.0
8	4/19	26	4	80.0	11	91.7	3	100.0	13	92.9	4	80.0
9	4/26	78	19	86.4	17	100.0	12	100.0	27	100.0	21	95.5
10	5/3	75	17	81.0	27	100.0	2	66.7	30	100.0	20	95.2
11	5/10	56	14	82.4	14	100.0	4	100.0	18	100.0	16	94.1
12	5/17	74	20	83.3	19	100.0	4	100.0	23	100.0	22	91.7
14	5/31	21	3	100.0	8	88.9	1	100.0	8	88.9	3	100.0
16	6/14	15	5	100.0	2	66.7	2	100.0	4	80.0	5	100.0
Total		**735**	**193**	**83.2**	**220**	**98.7**	**53**	**98.1**	**267**	**98.9**	**225**	**97.0**

Source: Spring 2010 sample management system.

Note: The sample size for the HOVRS observation and home visitor interview is 232. The sample size for the CLASS-T is 223. The sample size for the ITERS is 54. The sample size for the teacher interview is 267.

Ten teachers were observed for both the ITERS and CLASS-T because they had in their classes sample children from both cohorts.

Data collection did not occur during weeks 13 and 15.

Parents Who Exited Were Difficult to Reach

For the first round of the exit interviews, we completed 62 interviews (55 percent). We could not locate a working telephone number for 25 parents (22 percent) and could not reach or set a suitable interview time with 26 parents (23 percent).

In the second round of exit interviews, we completed 76 interviews for a response rate of 39 percent.[13] We could not locate a working telephone number for 65 parents (33 percent) and could not reach or set a suitable interview time with 45 parents (23 percent). Ten parents (5 percent) refused to complete the exit interview.

We attempted to conduct a total of 266 exit interviews across both rounds of data collection. Eight cases were ultimately ineligible for the exit interview because either the families returned to their Early Head Start study programs or the information we were provided turned out to be inaccurate and they had not left the programs in the first place. From the remaining 258 cases, we achieved a 54 percent response rate.

Table B.4. Exit Interview Response Rates

	Round One (Oct – Dec 2009)		Round Two (March – June 2010)		Overall	
	Number	Percentage	Number	Percentage	Number	Percentage
Complete	62	54.9	76	38.8	138	53.5
Not complete	51	45.1	120	61.2	120	46.5
Total	113	100.0	196	100.0	258	100.0

Source: Spring 2010 sample management system.

Note: The total number of cases released for round two includes the 51 incomplete responses from round one. Overall, we released a total of 258 unique cases.

Interviews Were Completed With All Program Directors for a Second Straight Year

We conducted telephone interviews with a program director or designee in all 89 programs (100 percent). In most cases (75 out of 89), the interview was conducted with the program director. In the other 14 cases, the OSC or another person designated by the program director completed the telephone interview. In 10 cases, more than one respondent participated in the telephone interview. We also received a total of 83 SAQs from program directors (93 percent).

Data Processing

Receipt Control Involved Several Steps and Validation Procedures

Receipt processes were the same as in spring 2009. After field materials returned to the SOC, SOC field staff reviewed the materials for each site, looking to ensure that all materials had arrived and that they matched the data collection plan.[14] If a child's teacher/home visitor had changed or

---

[13] The second round of exit interviews include (1) the exited parents we did not call during the first round because we did not receive a fall roster in time for data collection; (2) parents who exited after the first round; and (3) the 51 parents with whom we could not complete interviews during the first round.

[14] The field materials returned included teacher and home visitor interviews, HOVRS-A, ITERS-R, CLASS-T, SCRs, 24-month Child Assessment Record Forms, and parent SAQs.

was different than stated in the data collection plan, SOC field staff verified this change with the team leader and BFC and updated the SMS. After this review, SOC field staff transferred all materials to the receipting department at the SOC with a transmittal form to be scanned into the SMS. After being receipted, documents were placed into batches according to instrument type and transferred to quality control for review.

Quality Control and Data Entry use Two-Person Process

Quality control and editing of documents. We trained staff to review and edit documents for quality control (QC) and trained two QC supervisors to oversee the entire process. In February 2010, we trained three QC staff to review and edit the observation instruments (HOVRS-A, ITERS-R, CLASS-T). Shortly after, we trained three additional QC staff to review and edit the teacher/home visitor interviews, SCRs, and parent SAQs. In March 2010, two QC staff members were trained to review and edit the Child Assessment Record forms. QC staff reviewed the instruments for completeness and checked that the skip logic had been followed correctly. Where possible, QC staff made appropriate edits to the instruments based on pre-established specifications. The specifications dictated how QC staff should review scores, when to mark data as missing, and when to set data aside and flag it for review. Following review, project staff would instruct QC staff how to proceed.

Quality control of videos. Video recordings of the Two-Bag Task and ECI were captured on mini-DVDs that were transferred to the QC department after receipt. QC staff imported the videos onto PCs, added time stamps, and saved the video files to a secure drive for coding. If QC staff received a blank DVD or a DVD with no usable video (less than 3 percent of videos received), the video instruments were marked in the SMS as "Received Blank." At this stage, the QC supervisor created coding sheets from the SMS and assigned the videos to coders. The coders reviewed the videos and completed the appropriate coding sheet. If necessary, coders would place videos in supervisor review for reasons including poor lighting or audio, parents or children not being in the frame, and failure to meet the required length. However, videos were most commonly placed in supervisor review when an English coder was assigned a video including Spanish speakers. In each instance, the supervisor, in collaboration with project survey staff, reviewed the situation and provided the coder with directions to proceed. Completed coding sheets were placed into batches and transferred to data entry.

Data entry and coding. Project staff entered all data into the data entry program and coded responses. In 2010, we again used a two-person data entry process to ensure 100 percent verification. Project staff reviewed all verbatim and "other: specify" responses in the coding database. Staff back-coded responses into pre-existing answer options, built new codes if enough responses expressed the same concept, or left responses as verbatim text in the data file. After data were entered, the statuses of the instruments were updated in the SMS as "complete."

SMS and data file reconciliation. Staff conducted an extensive check to verify that the final instrument status in SMS matched what was in each data file. If completed instruments in a data file did not show up as complete in the SMS, staff investigated to determine the source of the discrepancy. In 2010, there were only a handful of discrepancies. A few were the result of a failure to mark instruments as "complete" after data entry. We also had a completed a parent interview and SCR for one exited case outside our data collection window; the case was removed from the final data files.

Two-Bag Task data file reconciliation. After data entry, the SAS programmer set up final video data files. Because the Parent-Child Interaction Rating Scales and PICCOLO are coded from the same video recording of the Two-Bag Task, project survey staff specifically reviewed these files to ensure they contained the same number of cases and that data such as language, gender, and length of video were consistent. After staff resolved any discrepancies, the data files were ready for frequency review.

Data Cleaning Consisted of Frequency Review and Data Editing

Frequency review. After data from all instruments were entered and considered complete, we ran frequencies for each data file. Survey staff responsible for each instrument reviewed the frequencies to verify that (1) the number of completed cases in the data file was correct; (2) the number of completed cases by cohort in the data file was correct; (3) the skip logic was followed correctly and each variable had the appropriate number of responses; (4) the frequency for each variable was feasible; (5) there were no missing data, additional data, or outliers; and (6) labels and variable names were correct.

Data editing. The staff responsible for each instrument made edits to the data when necessary after reviewing frequencies. The SAS programmer produced a spreadsheet for editing in which survey staff selected a variable to edit, entered the current value, entered the new value, and entered the reason why the value was being edited. A programmer read the specifications from these documents and updated the data file. All data edits were documented and saved in a designated file. Most data edits corrected minor data entry errors or interviewer/assessor coding errors identified during frequency review (for example, filled in missing data with "M" or cleared out "other: specify" verbatim data when the response had been back-coded). Each time a data file was updated, a new set of frequencies was run and reviewed. This process continued until all of the data files were clean and ready for analysis.

APPENDIX C. MEASURES

This appendix describes the measures used during spring 2010 data collection efforts to assess children in the 1-year-old Cohort at age 2. Given the longitudinal nature of Baby FACES, many of these measures were also used in baseline data collection.

We selected child and family outcome measures according to many considerations. Among our requirements were the reliability and validity of the measures, appropriateness for use with children and families from diverse backgrounds, comparability with other large research projects, burden on children and families, ease of administration and scoring, and appropriateness for use by Early Head Start programs. We also considered the need to complement well-established measures with those that are new to large-scale research and fill existing measurement gaps. We attempted to select measures that related to the cognitive, language, and social-emotional development outcome domains of the National Education Goals Panel.[15]

In addition to our review of the literature, we worked closely with experts from our technical work group (TWG), other authorities in the field, and the test developers themselves to select and modify measures for Baby FACES. The final list of measures presented here reflects the feedback of dozens of experts in the early childhood development field.

Measure Assessment and Scoring

We assessed the constructs arising from Baby FACES measures based on the user's guide for the measures or using a scoring approach consistent with the current literature. In addition, we used the following criteria in variable constructions:

- ***Sufficient Item-Level Data.*** If an individual was missing data from more than 25 percent of the items that made up a constructed variable, we did not compute a score for that individual. If the individual was missing up to 25 percent of the items, we imputed values based on the means of the items that were present. We used the specifications described in the user's guide to impute item values for the Ages and Stages Questionnaires, Third Edition (ASQ-3), and the Brief Infant Toddler Social Emotional Assessment (BITSEA). For methodological reasons, we did not impute missing data for the MacArthur-Bates Communicative Development Inventories (CDI)—Infant Short Form; the Preschool Language Scale, Fourth Edition (PLS-4); the Early Communication Indicator (ECI); the Home Visit Rating Scale-Adapted (HOVRS-A); the Infant Toddler Environment Rating Scale-Revised (ITERS-R); the Classroom Assessment Scoring System-Toddler (CLASS-T); and the Parent-Caregiver Relationship Scale (PCRS).[16]

- ***Adequate Internal Consistency Reliability.*** Methods of estimating reliability that require only a single test administration are referred to as measures of internal

---

[15] The National Education Goals Panel was commissioned by the president and governors of all 50 states in 1990. It established that all children should enter school ready to learn, and further described five domains of readiness that include (1) physical well-being and motor development, (2) social and emotional development, (3) approaches toward learning, (4) language development, and (5) cognition and general development.

[16] Because the HOVRS-A and ITERS-R scales are composed of means, we did not impute means for missing values. Similar considerations precluded imputing values for the CLASS-T and PCRS. The ECI score is based on occurrences of four communication elements, all of which need to be present to compute a total score. The raw scores for the CDI and PLS-4 are counts of correct answers, and the missing items are not counted for cases with 25 percent or fewer items missing. The scores are set to missing for cases with more than 25 percent of the items missing.

consistency or homogeneity. They are based on estimates of how well items within a scale or instrument measure the same cognitive domain or construct. We chose Cronbach's coefficient alpha, which captures the correlation among items on an assessment. The greater the covariance among items, the higher the reliability (and thus the higher the value of Cronbach's coefficient alpha). Values of the alpha can range from -1.0 to 1.0, with greater values indicating stronger internal consistency. Cronbach's coefficient alpha is an extension of Kuder-Richardson Formula 20, a measure of internal consistency that is used when the items are dichotomous (Cronbach 1951). We consider an alpha of 0.65 or higher as adequate for the constructed measures.

Psychometric Properties of Constructs

This section provides information on the selection criteria, normative samples, and psychometric properties reported by the developers of the measures for seven child outcome measures: the (1) ASQ-3 (Squires 2009), (2) MacArthur-Bates CDI (Fenson 2000), (3) PLS-4 (Zimmerman et al. 2002), (4) ECI (Luze et al. 2001; Carta et al. 2002), (5) BITSEA (Briggs-Gowan and Carter 2006), (6) Bayley Behavioral Rating Scale (BRS; Bayley 1993), and (7) child constructs derived from coding the parent-child play assessment using the Parent-Child Interaction Rating Scales for the Two-Bag Assessment (Mathematica Policy Research 2010). We also provide a brief background for other measures gathered during this wave of data collection, such as measures of parenting and the home environment, parent mental health, and home visit and classroom quality.

For the measures listed above, we analyzed the psychometric properties from the Baby FACES spring 2010 data collection and their difference sources. Tables C.1 through C.11 present the psychometric data for the constructed variables derived from the parent self-administered questionnaire (SAQ), staff-child report (SCR), direct child assessment, classroom and home visiting observations, and parent-child play assessment for the 1-year old Cohort at age 2. The tables are organized by measurement domain. We include the sample size, the possible range of values for each variable, the reported range in the Baby FACES sample, the unweighted sample mean, standard deviation, and the internal consistency reliability (coefficient alpha). Most of the constructed measures have internal consistency reliability of 0.65 or higher.

Measure of General Child Development

Ages and Stages Questionnaires, Third Edition. The ASQ-3 is a parent-report tool for screening infants and young children for developmental delays (Squires, Twombly, Bricker, and Potter 2009). The 21 questionnaires included in the ASQ-3 are appropriate for children ages 1 month to 5-1/2 years and focus on 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 rate questions such as "Does your child walk along furniture while holding on with only one hand?" on a scale of "not yet," "sometimes," or "most of the time." There are six items in each of the five developmental areas. The raw score in each developmental area could range from 0 to 60, and the ASQ-3 total area score could range from 0 to 300.

Due to the ASQ's widespread use by Early Head Start programs, we included it as a measure of a child's general development. Among the ASQ's advantages are its short administration time, psychometric soundness, relatively low cost, and availability in Spanish. The ASQ has demonstrated reliability, validity, and accuracy in distinguishing between children with and without developmental delays. Early Head Start programs often used this instrument to identify children with (or at risk for) development delays.

The normative sample includes 15,138 children between 1 month and 66 months of age throughout the United States. The sample includes more boys (53 percent) than girls (47 percent). Approximately two-thirds of children are white, 12 percent are African American, and 15 percent are Hispanic; other races make up the remaining 5 percent. More than half (54 percent) of mothers had at least four years of college, and only 3.5 percent had not completed high school. Most (57 percent) of the families have annual incomes greater than $40,000.

The psychometric studies on the ASQ-3 demonstrate adequate reliability and concurrent validity of the questionnaires. Intraclass correlations ranged from 0.75 to 0.82, indicating strong test-retest reliability across developmental domains. Inter-rater reliability is less strong; intraclass correlations by area range from 0.43 to 0.69. Cronbach's alphas range from 0.51 to 0.87. The ASQ classifications have moderate to high agreement with the Battelle Developmental Inventory (BDI) (Newborg 1984, 2004) classifications, with an aggregated sensitivity or specificity of 86 percent across all age intervals.

Cutoff points, which vary by age and indicate the need for further assessment, were derived by subtracting two standard deviations from the mean for each area of development (children scoring two standard deviations below the mean or lower are in the at-risk range). For example, the cutoff point in Communication is 13.04 for the 22-month form and 25.17 for the 24-month form. The cutoff point of two standard deviations has a sensitivity and specificity of 0.86. In other words, children 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 (between one and two standard deviations below the mean) might benefit from practicing skills in a specific area of development. As expected, the cutoff point of one standard deviation has a high sensitivity (0.98) but a low specificity (0.59). Therefore, some children who are developing normally will be classified as needing further assessment (Squires et al. 2009).

Table C.1 illustrates the average, standard deviation, range, and internal consistency of the ASQ-3 scores among children in the Baby FACES study. Cronbach's alphas for the study's sample are similar to previous studies.

Measures of Child Language Development

MacArthur-Bates Communicative Development Inventory. The CDI is designed to assess children's early receptive and expressive language and communication skills through parent report (Fenson et al. 2000). In the second year of data collection for Baby FACES, Early Head Start staff (teachers and home visitors) completed the English Toddler Short Form-Form A (a 100-word vocabulary checklist for 16- to 30-month-olds) for the 1-year-old Cohort children. Two measures were derived from this form:

- ***Vocabulary Comprehension*** measures the number of words the child understands. Teachers/home visitors are 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 report whether the child "understands and says" each of 100 specific words. The raw scores for both Vocabulary Comprehension and Vocabulary Production range from 0 to 100.

Teachers and home visitors who reported they spoke Spanish also completed the Spanish Toddler Short Form (a 100-word vocabulary checklist) for children identified as understanding Spanish.

Table C.1 Child General and Language Development at Age 2

Outcome	Possible Range		Reported Range		Mean/ Percentage	Standard Deviation	Cronbach Alpha
	Min.	Max.	Min.	Max.			
ASQ-3[a] Area Score							
Communication	0	60	5	60	48.49	12.62	0.75-0.77
Gross Motor	0	60	0	60	50.95	10.92	0.70-0.77
Fine Motor	0	60	5	60	44.34	10.77	0.50-0.64
Problem Solving	0	60	10	60	44.76	11.59	0.61-0.69
Personal-Social	0	60	5	60	49.85	10.31	0.63-0.68
Total Score	0	300	45	300	238.22	41.54	n.a.
ASQ-3[a] Total Scale Score							
Communication	0	110	10	110	83.40	23.49	n.a.
Gross Motor	0	90	0	90	69.06	16.67	n.a.
Fine Motor	0	110	20	110	73.18	19.43	n.a.
Problem Solving	0	130	20	130	93.33	22.67	n.a.
Personal-Social	0	110	10	110	86.70	18.31	n.a.
ASQ-3[a] IRT Score							
Communication	n.a.	n.a.	0	19	11.51	4.75	n.a.
Gross Motor	n.a.	n.a.	0	18	12.52	3.10	n.a.
Fine Motor	n.a.	n.a.	0	20	9.41	3.93	n.a.
Problem Solving	n.a.	n.a.	0	26	12.83	4.87	n.a.
Personal-Social	n.a.	n.a.	0	22	13.66	4.26	n.a.
ASQ Cut-Off Score (2SDs below the mean or lower)							
Communication	0	1	0	1	5.66	23.13	n.a.
Gross Motor	0	1	0	1	6.74	25.10	n.a.
Fine Motor	0	1	0	1	9.72	29.65	n.a.
Problem Solving	0	1	0	1	9.59	29.47	n.a.
Personal-Social	0	1	0	1	6.58	24.81	n.a.
ASQ in the Monitoring Zone (1 to 2SDs below the mean)							
Communication	0	1	0	1	11.32	31.71	n.a.
Gross Motor	0	1	0	1	11.24	31.61	n.a.
Fine Motor	0	1	0	1	20.19	40.18	n.a.
Problem Solving	0	1	0	1	12.78	33.42	n.a.
Personal-Social	0	1	0	1	10.15	30.23	n.a.
Staff-Reported CDI[b] (English) Raw Score							
Vocabulary Comprehension	0	100	0	100	69.06	23.50	0.98
Vocabulary Production	0	100	0	100	35.06	26.02	0.99
Staff-Reported CDI[b] (English) IRT Score	n.a.	n.a.	-10	9	0.35	2.72	n.a.
Staff-Reported CDI[b] (Spanish) Raw Score	n.a.	n.a.	n.a.	n.a.			
Vocabulary Comprehension	0	100	6	100	71.10	22.81	0.98

Outcome	Possible Range		Reported Range		Mean/ Percentage	Standard Deviation	Cronbach Alpha
	Min.	Max.	Min.	Max.			
Vocabulary Production	0	100	0	98	30.84	25.74	0.99
Parent-Reported CDI[a] English							
Vocabulary Production	0	100	0	100	47.07	25.62	0.98
Parent-Reported CDI[a] Spanish							
Vocabulary Production	0	100	0	100	46.89	27.75	0.98
PLS-4[c] Standard Score							
English	50	150	50	139	90.76	16.77	0.88
Spanish	50	150	51	148	91.06	21.13	0.89
Bilingual	50	150	52	148	95.28	20.11	0.89
ECI – Expressive Language							
ECI Standard Score	n.a.	n.a.	62	158	91.00	18.59	n.a.[d]
ECI language delay (percentage 1.5 SDs below the mean or lower)	0	1	0	1	29.40	45.60	n.a.
ECI at-risk for language delay (percentage 1 to 1.5 SDs below the mean)	0	1	0	1	12.60	33.22	n.a.

Sample Size
Parent Interview 107-537
PLS-4 English 385
PLS-4 Spanish 140
PLS-4 Bilingual 144
SCR English CDI 551
SCR Spanish CDI 102
ECI 500

Source: Spring 2010 Parent Interview, Staff-Child Report, Direct Child Assessment, and Parent-Child Play Interaction (Two-Bag Task).

Note: Sample restricted to the 1-year-old Cohort. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 2.

[a] Parent report.
[b] Teacher/home visitor report.
[c] Direct Child Assessment.
[d] ECI scores are based on weighted counts of children's use of gestures, vocalizations, and single- and multiple-words; these are not conceptually correlated. Cronbach's alpha, or the extent to which a scale's individual items tap the same underlying construct, is thus not an appropriate metric for assessing reliability. We instead report estimates of inter-rater reliability.

n.a. = not applicable; ASQ-3 = Ages & Stages Questionnaires (Third Edition); CDI = MacArthur Communicative Development Inventories; PLS-4 = Preschool Language Scale (Fourth Edition); ECI = Early Communication Indicator; SCR=Staff Child Report. Sample restricted to 1-year-old Cohort at age 2.

In addition to staff reports, parents of 1-year-old Cohort children also completed the English or Spanish Toddler Short Form for children's Vocabulary Production in the self-administered questionnaire.

The CDI was used successfully in the Early Head Start Research and Evaluation Project (EHSREP) despite concerns about the norming sample's appropriateness. EHSREP researchers found that Early Head Start had a significant positive impact on 24-month-old children's language production. All versions of the CDI also show concurrent validity with other measures such as the Bayley language subscales. The ability to have both parents and home visitors provide data on this instrument made this measure of language development in Baby FACES a valuable tool.

The norming sample for the English Toddler Short Form includes 898 toddlers between 16 and 30 months of age from three locations in the United States: New Haven, Connecticut; Seattle, Washington; and San Diego, California. The majority (93 percent) of children are white. Black and Asian children make up 1 and 2 percent of the sample, respectively. The remaining 5 percent is made up of other races. More than half (54 percent) of parents hold a college diploma, and only 2 percent have not completed high school. The upwardly skewed socioeconomic status (SES) distribution of the normative sample may limit the applicability of the norms to children from low-SES families. The normative sample was also limited to children whose primary language was English. Approximately 13 percent of the toddlers in the sample had exposure to more than one language. The normative study for the Spanish Toddler Form is still under review.

Cronbach's alpha was 0.99 for the toddler Form A in the normative sample. The correlation between the short and long toddler forms is 0.74 for Form A, suggesting that the short form provides an effective alternative to the long form.

For 1-year-old Cohort children at year two of data collection, parent- and staff-reported vocabulary production scores are moderately correlated ($r = 0.49$ for English and 0.38 for Spanish).

Table C.1 illustrates the average, standard deviation, range, and internal consistency of the CDI scores among children in the Baby FACES study.

Preschool Language Scale, Fourth Edition (PLS-4). The PLS-4 is a direct child assessment used to evaluate children's receptive and expressive language skills, as well as information about children's understanding and use of grammatical rules from birth through 7 years of age (Zimmerman et al. 2002). It comprises two subscales: (1) Auditory Comprehension (AC) and (2) Expressive Communication (EC). We used the AC subscale for both the English and Spanish editions of the PLS-4 during the second year of Baby FACES data collection. The tasks designed for infants and toddlers in the AC subscale assess skills that are important for language development (for example, following directions with cues and appropriate object play).

For Spanish-speaking children, we also administered in English the test items that they did not pass in the Spanish version until they reached the ceiling for the English test. Thus, we were able to calculate the conceptual scores by crediting children 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 standardization sample for the PLS-4 English included 1,564 children ages 2 days to 6 years, 11 months (about 100 children for each age group from 18 months to 36 months). A stratified representative sample was selected based on parent education level, geographic area, and race. About one-third of the sample was from the South; children from the West and North Central each made up a

quarter of the sample; and 18 percent of the sample were from the Northeast. Within each age level, the proportion of boys and girls was split evenly. About 17 percent of the sampled children's primary caregivers were without a high school diploma; 32 percent of primary caregivers had a high school diploma or GED; 28 percent had earned some college credit; and 23 percent had completed four or more years of college. Most (97 percent) of the children spoke English only, and 3 percent spoke languages other than English.

The standardization sample for the PLS-4 Spanish included 1,188 children ages 2 days to 6 years, 11 months. A stratified representative sample was selected based on parent education level and geographic area. About half of the sampled children were from the South; 43 percent from the West; and the remaining 7 percent from the Northeast and North Central. The proportion of boys and girls was split evenly within each age level in the sample. About 62 percent of the sample was white; 15 percent African American; 17 percent Hispanic; and 5 percent comprising other races. Slightly less than half of the sample included children whose primary caregivers were without a high school diploma; 22 percent of primary caregivers had a high school diploma or GED; 14 percent had spent one to three years in college or a technical school; and 17 percent had completed four or more years of college. Almost all children in the sample were Hispanic, and most of them (81 percent) spoke a Spanish dialect used in Mexico.

The psychometric evidence indicates that the PLS-4 English and Spanish can provide reliable and valid inference about a child's language ability (Zimmerman et al. 2002). For the PLS-4 English, the internal consistency reliability coefficients for the AC subscale range from 0.91 to 0.94 for children ages 18 months to 41 months. The test-retest reliability estimates range from 0.87 to 0.95 for children ages 24 months to 41 months. The PLS-4 correctly identified 79 percent of 3-year-old children with a previously diagnosed language disorder and 92 percent of typically developing 3-year-old children.

For the PLS-4 Spanish, the internal consistency coefficients for the AC subscale range from 0.84 to 0.89 for children ages 18 to 42 months. The test-retest reliability estimates for children ages 24 months to 47 months range from 0.73 to 0.84. PLS-4 correctly identified 87 percent of children with a previously diagnosed language disorder and 57 percent of typically-developing 3-year-old children.

Table C.1 illustrates the average, standard deviation, range, and internal consistency of the PLS-4 scores among children in the Baby FACES study.

Early Communication Indicator. As part of the assessment activities conducted with 2-year-olds, interviewers administered the ECI–a semi-structured, play-based communication task designed to measure the expressive communication of infants and toddlers between the ages of 6 months and 36 months (Luze et al. 2001; Carta et al. 2002, 2010). The ECI comprises four key skill elements, or communicative behaviors:

1. ***Gestures*** are nonverbal, intentional actions that convey communicative intent (such as pointing to direct attention to an object).

2. ***Vocalizations*** consist of nonword, verbal utterances (such as cooing or babbling), or verbalizations voiced by the child that are otherwise unintelligible.

3. ***Single-word utterances*** are defined as single words voiced by the child that are recognizable and readily understood (such as "pig" and "bye-bye").

4. ***Multiple-word utterances*** consist of two or more voiced words that fit together in a meaningful way to approximate a statement or sentence (such as "Piggy sleeps" and "Cow eats food").

Play interactions were video-recorded for subsequent coding by staff at Mathematica. Coders record the frequency of occurrences of each observed skill element over the six-minute assessment; observed instances are later combined to yield a total communication score. Specifically, the total communication score 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.[17] The total weighted score is then converted to a rate score that reflects the number of communicative acts per minute over the course of the six-minute play assessment. An age-based, standardized score with a mean of 100 (SD = 15) is also calculated. 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).

Beyond providing a measure of children's communication proficiency at discrete periods during early development, the ECI can also be administered on an ongoing basis to monitor the short-term growth and development of children's expressive communication over time (Carta et al. 2002; Luze et al. 2001). Other notable advantages include its short administration time, ease of administration, and psychometric soundness.

The feasibility and psychometric properties of the ECI have been documented across a number of studies, including a longitudinal study of 50 children in center-based care (Luze et al. 2001); a cross-sectional study of 1,486 infants and toddlers served by Early Head Start, community-based childcare, and early intervention programs (Greenwood et al. 2006); and, most recently, a large-scale study of 5,883 children enrolled in Early Head Start programs across two states. The broader goal of the large-scale study was to develop a normative sample based on children served by Early Head Start (Greenwood et al. 2010).

Normative growth estimates are based on a composite sample of more than 1,400 children that combined data from three smaller study samples (Greenwood et al. 2006). This aggregate sample represents children drawn from 22 center- and 14 home-based programs from 1999 to 2004. Most (90 percent) of these children were participating in Early Head Start. Collectively, children were racially diverse, with 49 percent from African American, Hispanic, or racially mixed minority backgrounds. Slightly more than half of all children (55 percent) were male, and 12 percent were receiving Part C early intervention services with Individualized Family Service Plans. Ten percent of families reported languages other than English were spoken in the home. Although the combined sample was predominantly low-income, it did include some children from middle- to high-SES families. The total weighted communication score for 2-year-old children in the sample averaged 11.8. Normative estimates derived from this sample were used in the creation of ECI total communication standard scores in the Baby FACES study.

A more recent, population-specific normative sample is based on data obtained between 2002 and 2007 from 5,883 children served by Early Head Start in two midwestern states (Greenwood et al. 2010). However, normative estimates by child age (in months) are currently unavailable. Assessments were obtained at six-month intervals between 6 months and 36 months, with 18.4 months as the average age at the first assessment. Overall, the median number of observations per child was three,

---

[17] Weighting is used in the calculation of the total communication score to create a growth-based metric that reflects growth in communication proficiency by offsetting declines in prelinguistic communication (gestures and vocalizations) that occur as children acquire greater proficiency in spoken language skills. The weighted calculation also approximates an absolute estimate of total words produced by the child.

and the mean number of assessments obtained at each month of age was 428 (SD = 198; range = 23 to 727). Children were regionally and racially diverse, with nearly all children residing in homes in which English (90 percent) or Spanish (9 percent) were spoken. Representation by gender was roughly equivalent (48 percent male). In accordance with Early Head Start mandates, 8 percent of children were receiving Part C services. Among 2-year-old children in the sample, the total weighted communication score averaged 13.0.

Studies of the ECI demonstrate that it can be reliably administered, with reported split-half reliabilities of 0.89 and interobserver agreement of 90 percent (Luze et al. 2001; Greenwood et al. 2004). The ECI has also been shown to demonstrate significant, positive associations with known measures of early communication, including the Preschool Language Scale, Third Edition (PLS-3; Zimmerman et al. 1992) and the Caregiver Communication Measure (CCM; Walker et al. 1998; r = 0.62 and 0.51, respectively). Cross-sectional and longitudinal studies have documented the ECI's sensitivity to individual differences in communication proficiency within age and across early development, sensitivity to differences in the performance of children with disabilities, and sensitivity to short-term early interventions with infants and toddlers in Early Head Start (Carta et al. 2004; Greenwood et al. 2002, 2006; Luze et al. 2001).

Inter-rater reliabilities between the team leaders and coders were established to a criterion of 80 percent agreement. Thereafter, the team conducted weekly inter-rater reliability checks on three to five randomly selected videos. A total of 54 videos (10.7 percent of the 506 codable videos) served as reliability videos.[18] Agreement averaged 82 percent across all coders, with a range of 78 to 85 percent.

Table C.1 illustrates the average, standard deviation, and range of the ECI scores among children in the Baby FACES study.

Correlations Among Language Measures Used in Baby FACES. To compare relationships and test the validity of the language measures and sources of report, we ran correlations of all language measures for the 1-year-old Cohort at age 2 (Table C.2).

The correlations between the CDI parent-reported English and Spanish vocabulary production scores and the ASQ-3 Communication IRT score are both higher than 0.60 (0.65 for English and 0.62 for Spanish). The correlations of the CDI staff-reported vocabulary production scores with the ASQ-3 IRT score are lower than parent reports (0.44 for English and 0.27 for Spanish). The ASQ-3 Communication score is correlated with the CDI staff-reported English vocabulary comprehension score (0.30) but is not correlated with the CDI staff-reported Spanish vocabulary comprehension score (0.06).

---

[18] We received a total of 519 videos of ECI administrations from the field. We were unable to code one video due to poor audio quality. An additional 12 videos were shorter than the required duration and were excluded from the analyses. Therefore, we computed scores from a total of 506 videos.

Table C.2. Correlations Among Language Measures at Age 2

	ASQ-3 Communication IRT score	Parent-Reported CDI English Production	Parent-Reported CDI Spanish Production	Staff-Reported CDI English Production	Staff-Reported CDI Spanish Production	Staff-Reported CDI English Comprehension	Staff-Reported CDI Spanish Comprehension	PLS-4 English	PLS-4 Spanish	PLS-4 Bilingual
Parent-Reported CDI										
English Vocabulary Production	0.65***	.	.	.	.	.	.	.	.	.
Spanish Vocabulary Production	0.62***	n.a	.	.	.	.	.	.	.	.
Staff-Reported CDI										
English Vocabulary Production	0.44***	0.49***	0.19+	.	.	.	.	.	.	.
Spanish Vocabulary Production	0.27**	-0.10	0.38***	0.40***	.	.	.	.	.	.
English Vocabulary Comprehension	0.30***	0.31***	0.08	0.64***	0.14	.	.	.	.	.
Spanish Vocabulary Comprehension	0.06	0.03	0.22+	0.00	0.65***	0.02	.	.	.	.
PLS-4										
English	0.37***	0.36***	n.a	0.36***	0.08	0.23***	-0.44	.	.	.
Spanish	0.27**	-0.17	0.31**	-0.19*	0.00	-0.24**	0.16	n.a.	.	.
Bilingual	0.30***	0.21	0.28**	-0.10	0.03	-0.17+	0.11	n.a.	0.91***	.
ECI Total Communication	0.26***	0.33***	0.22*	0.26***	0.24*	0.12*	0.10	0.15**	0.12	0.11

Source: Spring 2009 Parent Interview, Staff Child Report (SCR) Direct Child Assessment, and Parent-Child Play Interaction (Two-Bag Task).
Note: Sample restricted to 1-year-old Cohort.

n.a.= not applicable; ASQ-3 = Ages & Stages Questionnaires (Third Edition); CDI = MacArthur Communicative Development Inventories; PLS-4 = Preschool Language Scale (Fourth Edition); ECI = Early Communication Indicator.

+$p < 0.10$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

The correlations of the CDI parent- and staff-reported English vocabulary production with the PLS-4 English scores are both 0.36. Both the CDI parent- and staff-reported English production scores are negatively correlated the PLS-4 Spanish scores, although the correlations are low (-0.17 to -0.19). The CDI parent-reported Spanish production score is correlated with the PLS-4 Spanish score and bilingual score at 0.31 and 0.28, respectively. However, the correlations of the CDI staff-reported Spanish production or comprehension score with the PLS-4 Spanish score or bilingual score are low to none.

The correlations between parent- and staff-reported CDI English and Spanish scores and the ECI total communication are in the low range (0.22 to 0.33). The correlations of the ASQ-3 Communication with the PLS-4 scores and the ECI total communication are in the low to moderate range (0.26 to 0.37). The correlations of the PLS-4 scores with the ECI total communication are low (0.11 to 0.15).

Measures of Child Social-Emotional Development

Brief Infant Toddler Social Emotional Assessment. The BITSEA is the screener version of the longer Infant Toddler Social Emotional Assessment (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 months to 36 months old (Briggs-Gowan and Carter 2006). 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).

We selected the BITSEA as our measure of social-emotional development due to its dual focus on both social competencies and behavior problems, such as internalizing and externalizing behavior, Spanish language availability, and the possibility of administering it to both parents and staff.

The 31-item BITSEA Problem scale assesses social-emotional/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 less 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 we administered it to both parents and teachers/home visitors in second year of data collection. The raw score ranges from 0 to 22 for the competence domain and 0 to 62 for the problem domain.

We created cutoff scores to indicate a high degree of problems or low competence. We calculated cutoff points in six-month age bands according to child gender by using cutoff points established with the national standardization sample. For the BITSEA Problem scale, the cutoff point indicates scores at the 75th percentile or higher. For the BITSEA Competence scale, the cutoff point indicates scores at the 15th percentile or lower. A score in this range suggests that delays in social-emotional competence may be present. Scoring in the cutoff range in either or both domains (that is, many problems and/or low competence) indicates "screening positive" on the BITSEA.

The nationally normative sample includes 600 children between 12 months and 35 months, 30 days of age, with 150 children (75 boys and 75 girls) in each age band: 12 to 17 months, 18 to 23 months, 24 to 29 months, and 30 to 35 months. Each age band was stratified to match the 2002 U.S. Census on race/ethnicity, parent education level, and region.

The BITSEA has adequate test-retest reliability ($r = 0.82$–0.92), inter-rater reliability (r = 0.67–0.74) (Briggs-Gowan and Carter 2006), and internal consistency (Cronbach's alpha of 0.79 for the Problem scale and 0.65 for the Competence scale on the Parent Form, as well as Cronbach's alpha of 0.80 for the Problem scale and 0.66 for the Competence scale on the Childcare Provider Form) (Briggs-Gowan 2004).

The BITSEA has demonstrated construct validity through expected associations with other measures of the same construct (Briggs-Gowan and Carter 2006). The BITSEA Parent Form Problem and Competence scores were both moderately correlated with the ASQ: Social-Emotional (ASQ: SE) (Squires 2002) (r = 0.55 and r = -0.55, respectively). The correlations between the BITSEA Problem score and the Child Behavior Checklist 1.5–5 (CBCL 1.5–5) (Achenbach 2000) Internalizing, Externalizing, and total scores range from 0.46 to 0.60, and the correlations between the BITSEA Competence score and the CBCL scores range from -0.30 to -0.42. The BITSEA scores were moderately correlated with the Adaptive Behavior Assessment System—Second Edition (ABAS-II) (Harrison 2003) domain-specific skill scores (Conceptual, Social, and Practical), with the correlations ranging from 0.39 to 0.56 for Competence and -0.31 to -0.36 for Problem. The BITSEA scores also demonstrated small to modest correlations with the Bayley Scales of Infant and Toddler Development—Third Edition (Bayley-III) (Bayley 2006) Cognitive Assessment and Language Scale, ranging from 0.25 to 0.32 for Competence and -0.19 to -0.28 for Problem. The correlation between the BITSEA Problem score and the Bayley-III Social Emotional score was -0.27, and the correlation between the BITSEA Competence score and the Bayley Social-Emotional score was 0.51.

The BITSEA has also demonstrated validity in discriminating children with clinically significant problems from matched control subjects (Briggs-Gowan and Carter 2006). The BITSEA Competence scale demonstrates excellent sensitivity (100 percent) and good specificity (91 percent) in detecting autistic disorder. The Problem scale provides excellent specificity (97 percent) and some sensitivity (64 percent).

The BITSEA validation study (Briggs-Gowan 2004) reported that the parent and child care provider correlation was higher than expected for Competence (0.59) and typical for Problems (0.28), because children may behave differently in the two contexts. For the second year of Baby FACES data collection, the correlations between parent ratings and Early Head Start staff ratings are lower than those found in the BITSEA validation study (r = 0.31 for Competence and 0.15 for Problem). Although still low, for the Problem scale, home visitor ratings are more highly correlated with parent ratings ($r = 0.22$) than are teacher ratings with the parent ratings ($r = 0.13$) on the Problem scale; for the Competence scale, teacher ratings are more highly correlated with parent ratings ($r = 0.35$) than are home visitor ratings and parent ratings ($r = 0.27$).

Table C.3 illustrates the average, standard deviation, range, and internal consistency of the BITSEA scores among children in the Baby FACES study.

Bayley Behavioral Rating Scale (BRS). The BRS measures a 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). Baby FACES uses the following two subscales:

- ***Orientation/Engagement*** measures the child's cooperation with the interviewer 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.

Table C.3 Child Social-Emotional Development at Age 2

Outcome	Possible Range		Reported Range		Mean/ Percentage	Standard Deviation	Cronbach Alpha
	Min.	Max.	Min.	Max.			
Parent-Reported SAQ BITSEA Raw Score							
Problem Domain	0	62	0	48	12.20	7.53	0.87
Competence Domain	0	22	1	22	16.93	3.43	0.76
Staff-Reported BITSEA Raw Score							
Problem Domain	0	62	0	28	7.71	5.76	0.82
Competence Domain	0	22	1	22	15.72	3.67	0.80
Parent-Reported SAQ BITSEA Cut-Off Score							
Problem Domain	0	1	0	1	35.47	47.89	n.a.
Competence Domain	0	1	0	1	20.97	40.75	n.a.
Staff-Reported BITSEA Cut-Off Score							
Problem Domain	0	1	0	1	24.16	42.85	n.a.
Competence Domain	0	1	0	1	19.67	39.78	n.a.
Parent-Reported SAQ BITSEA Screen Positive	0	1	0	1	45.01	49.80	n.a.
Staff-Reported BITSEA Screen Positive	0	1	0	1	36.92	48.30	n.a.
Assessor-Reported BRS Total Scale Score							
Orientation/Engagement			9	45	32.50	7.50	0.92
Emotional Regulation			10	50	36.37	8.70	0.92
Assessor-Reported BRS in Questionable Range							
Orientation/Engagement	0	1	0	1	21.85	41.36	n.a.
Emotional Regulation	0	1	0	1	22.28	41.65	n.a.
Assessor-Reported BRS in Non-Optimal Range							
Orientation/Engagement	0	1	0	1	26.67	44.26	n.a.
Emotional Regulation	0	1	0	1	43.46	49.62	n.a.
Parent-Child Interaction Rating Scales							
Engagement of Parent	1	7	2	7	4.39	1.19	n.a.
Sustained Attention	1	7	1	7	4.88	1.02	n.a.
Negativity Toward Parent	1	7	1	7	2.81	1.37	n.a.
Enthusiasm	1	7	1	7	4.39	1.08	n.a.
Sample Size							
Parent Interview			**511-516**				
SCR			**538-557**				
Assessor Rating			**540-543**				
Parent-Child Play Interaction			**511**				

Source: Spring 2010 Parent Interview, Staff-Child Report, Direct Child Assessment, and Parent-Child Play Interaction (Two-Bag Task).

Note:Sample restricted to the 1-year-old Cohort. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 2.

n.a.= not applicable; BITSEA = Brief Infant-Toddler Social & Emotional Assessment; SCR=Staff Child Report; BRS = Bayley II Behavior Rating Scale.

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 test-retest reliability coefficients range from 0.60 to 0.71 for ages 24 months and 36 months. The correlations between the two subscales and the Bayley Scales of Infant Development Mental Development Index (MDI) and Bayley Scales of Infant Development Psychomotor Development Index (PDI) are in the low range (0.22 to 0.34), suggesting that the BRS taps a unique source of variance from the Mental and Motor Scales. The BRS "nonoptimal" cutoff scores indicate raw scores at or below the 10th percentile, and "questionable" cutoff scores indicate raw scores between the 11th and 25th percentiles. The BRS cutoff scores differentiate children with significant impairment (identified by the MDI and PDI or diagnosed with medical conditions that are associated with severe impairment) from normally developed children.

Table C.3 illustrates the average, standard deviation, range, and internal consistency of BRS scores among children in the Baby FACES study.

Parent-Child Interaction Rating Scales for the Two-Bag Assessment–Child Scales. As part of the assessment activities conducted with 2-year-olds, we administered an eight-minute semi-structured play-based task to parents and children (Two-Bag Task). Child behaviors were assessed using the Parent-Child Interaction Rating Scales for the Two-Bag Assessment (Mathematica Policy Research 2010). Collectively, the 12 rating scales assess a range of child and parent behaviors. Nine of the component scales are derived from the 24- and 36-Month Child-Parent Interaction Rating Scales for the Three-Bag Assessment that was used as part of the EHSREP (ACF 2002; Brady-Smith et al. 1999, 2000). Three additional scales that originated in the NICHD Study of Early Child Care Three Box Task (NICHD Early Child Care Research Network 1997, 1999) were adapted from Cox (1997). The coding scheme includes six scales that focus on parent behaviors; four scales that assess child behaviors; and two scales that address the quality of the dyadic interaction as a whole. Each area is assessed on a seven-point scale, ranging from a very low incidence of the behavior to a very high incidence of the behavior. Ratings along the scale are anchored by a description of the behaviors (and associated exemplars) that warrant a specific score. Overall, the scales measure both the prevalence and intensity of the observed behaviors.

Four scales assess the child's (1) engagement of parent (extent to which the child initiates and/or maintains interaction with the parent); (2) sustained attention with objects (degree of involvement with and focused exploration of the play materials); (3) enthusiasm (degree of vigor and confidence during the task); and (4) negativity toward parent (displays of anger, hostility, or disdain). (See Parent-Child Interaction Rating Scales for the Two-Bag Assessment–Parent Scales for information on the eight parenting-related scales, including psychometric properties). Based on findings from preliminary analyses of the Baby FACES data, we retained the child scales as individual scales. The scales were significantly intercorrelated, with associations that were moderate to high in magnitude (-0.43 to 0.73).

The psychometric properties of the scales have been documented in other large-scale studies, including the EHSREP and ECLS-B.[19] Interobserver agreement (exact or within one point) on the

---

[19] In both of these studies, the parent-child play assessment was completed by parents and children during visits conducted when children were 2 years of age (as one of a number of repeated assessments). However, the task on which coding was based varied slightly across the studies. Specifically, in the EHSREP, parents and children engaged in play with materials provided in three numbered bags (Three-Bag task): a children's book, a cooking set, and a Noah's Ark set

national coding scales in the EHSREP averaged 90 percent at 14 months, 93 percent at 24 months, and 94 percent at 36 months (with a range of 83 percent to 100 percent across the three ages; Brady-Smith et al. 2005). In ECLS-B, interobserver agreement on the child scales at age 2 was 94.7 percent (Andreassen et al. 2007). The scales have also been shown to relate significantly to widely used instruments that tap similar child constructs (Ispa et al. 2002). For example, mothers' concurrent ratings of their children's behavior on the Aggressive subscale of the Child Behavior Checklist (CBCL; Achenbach et al. 1987) has been shown to correlate positively with ratings of child negativity ($r = 0.16$) and negatively with ratings of child engagement ($r = -0.19$).

A total of 55 videos (10.8 percent of the 507 codable videos) served as reliability videos.[20] Agreement (exact or within one point) averaged 86 percent across all coders, with a range of 85 to 89 percent. Cohen's kappa[21] for scales assessing child behaviors ranged from 0.66 to 0.77.

Table C.3 illustrates the average, standard deviation, range, and internal consistency of scores for the child constructs derived from the Parent-Child Interaction Rating Scales in Baby FACES.

Measures of Positive and Negative Parenting Behaviors During Play

Ratings of parent behaviors were based on coding of the Two-Bag Task according to 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).

Parent-Child Interaction Rating Scales for the Two-Bag Assessment–Parent Scales. As described above, the Parent-Child Interaction Rating Scales (Mathematica Policy Research 2010) comprises 12 scales that assess a range of child and parent behaviors, eight of which are related to parenting. Six scales focus on specific parent behaviors, and two scales more broadly address the quality of the dyadic interaction. (See Parent-Child Interaction Rating Scales for the Two-Bag Assessment–Child Scales for more information on the development of the measure, and for psychometric information on the Child Scales.)

Collectively, the parenting and dyadic scales address: (1) sensitivity (the extent to which the parent acknowledges the child's perspective, accurately perceives the child's signals, and promptly and appropriately responds to these signals); (2) positive regard (displays of love, respect, and/or admiration); (3) stimulation of cognitive development (effortful teaching aimed at expanding the child's abilities); (4) intrusiveness (over-involvement and over-control); (5) detachment (under-involvement with, lack of awareness of, attention to, and engagement of the child); (6) negative regard (expressions of discontent with, anger toward, and rejection of the child); (7) relationship

---

with various animals. In ECLS-B, the parent and child dyad were asked to play with materials provided in two numbered bags (Two-Bag Task): a set of toy dishes and a children's book.

[20] We received a total of 522 videos of Two-Bag administrations from the field. We were unable to code 11 using the Parent-Child Interaction Rating Scales for a number of reasons, including a parent speaking a language other than English or Spanish ($n = 7$), poor audio quality ($n = 2$), a parent and/or child being out of frame ($n = 1$), and an interaction that was not recorded due to field interview error ($n = 1$). An additional four videos were shorter than the required duration and were excluded from the analyses. Therefore, we computed scores from a total of 507 videos.

[21] Coefficients were weighted to reflect the degree of disagreement among coder ratings. Whereas unweighted kappa treats all disagreements equally, weighted kappa attaches greater emphasis to large differences between ratings along the ordinal scale. For example, disagreement by one scale point is seen as less serious than disagreement by two scale points, and so on.

quality (degree of relatedness and mutual engagement); and (8) boundary dissolution (extent to which the parent fails to maintain an appropriate parental role in his or her interaction with the child).

The psychometric properties of the scales have been documented in other large-scale studies, including the EHSREP and ECLS-B (see Parent-Child Interaction Rating Scales for the Two-Bag Assessment–Child Scales for interobserver agreement estimates in the EHSREP study sample). In ECLS-B, interobserver agreement at age 2 on the parent scales was 96.5 percent (Andreassen et al. 2007). Cronbach's alpha for the composite measure of supportive parenting created in the EHSREP—derived from average scores on sensitivity, cognitive stimulation, and positive regard, all of which were strongly intercorrelated (rs = 0.50 to 0.71)—ranged from 0.82 to 0.83 at the three ages (Brady-Smith et al. 2005). A composite derived from these three scales was also created for the ECLS-B age 2 data collection, although detailed psychometric information is not reported.

The parent rating scales have also been shown to relate significantly to widely used instruments that tap similar parent constructs (Ispa et al. 2002). Positive associations were demonstrated between intrusiveness ratings and scores on the Traditional subscale ($r = 0.22$) of the Parental Modernity Scale (Schaefer and Edgerton 1985), and ratings of parents' positive regard and the Emotional Responsivity subscale ($r = 0.30$) of the Infant/Toddler Home Observation for Measurement of the Environment (HOME; Caldwell and Bradley, 1984). Negative associations between dyadic mutuality ratings and mothers' concurrent scores on the Parent-Child Dysfunctional Interaction subscale ($r = -0.17$) of the Parenting Stress Index (PSI; Abidin 1995) have also been reported.

Cohen's kappa[22] for scales assessing parent behaviors ranged from 0.42 to 0.59; kappa coefficients for relationship quality and boundary dissolution were 0.66 and 0.41, respectively.

We conducted preliminary analyses of the Baby FACES data to examine patterns of association among the scales, possible underlying factors, and internal consistency reliability. Based on our analyses, we created a composite parenting score, "synchronicity" (Cronbach's alpha = 0.88), by computing a mean score derived from scores on parental sensitivity, positive regard, and relationship quality—all of which were highly and significantly correlated (ranged from 0.63 to 0.79).[23] We retained the scales assessing negative parenting behaviors (intrusiveness, detachment, negative regard, and boundary dissolution) as individual scales. The correlations among the four negative parenting scales were small to moderate and statistically significant (0.12 to 0.40), with the exception of associations between negative regard and intrusiveness and boundary dissolution (0.55 and 0.46, respectively).

---

[22] Coefficients were weighted to reflect the degree of disagreement among coder ratings. Whereas unweighted kappa treats all disagreements equally, weighted kappa attaches greater emphasis to large differences between ratings along the ordinal scale. For example, disagreement by one scale point is seen as less serious than disagreement by two scale points, and so on.

[23] To allow for comparisons to other large-scale studies such as EHSREP, a second composite score of positive parenting, "supportiveness," (Cronbach's alpha = 0.73) was derived from scores on parental sensitivity, positive regard, and cognitive stimulation (correlations ranged from 0.32 to 0.69). Notably, associations between parental cognitive stimulation and parental sensitivity ($r = 0.32$), positive regard ($r = 0.42$), and relationship quality ($r = 0.35$) were only moderate in magnitude, and inclusion of cognitive stimulation in the composite score reduced the alpha from 0.88 to 0.82. Given the overall lower internal consistency reliability and intercorrelation of its components, we reserve discussion of this construct for anchoring our findings to those reported in other national studies. Of note, concurrent associations between key child development outcomes and synchronicity were consistently more robust than were associations to supportiveness, lending further support to the validity of this composite measure.

Parenting Interactions with Children: Checklist of Observations Linked to Outcomes. The PICCOLO is an observational instrument designed to measure "developmental parenting" along four key domains known to support children's development across a number of areas (Cook and Roggman 2009; Roggman et al. 2009). The measure was developed for use by practitioners working with parents of young children and has applications to other settings, including research and intervention efforts. Specifically, the PICCOLO rates 29 positive parenting behaviors along four domains: (1) Affection (displays of warmth, physical closeness, and positive expressions toward the child); (2) Responsiveness (to the child's cues, emotions, vocalizations, interests, and behaviors); (3) Encouragement (attempts to support the child's exploration, effort, skills, initiative, curiosity, creativity, and play); and (4) Teaching (the degree to which the parent engages in shared conversation and play, provides cognitive stimulation, and extends the child's verbalizations). All items are rated on a three-point scale, ranging from behaviors that are "absent" or not observed (0) to those that are "clearly evident" and frequent in their occurrence and/or intensity (2). A score of 1 indicates emerging behaviors that are "barely" or briefly observed. The domains of Affection, Responsiveness, and Encouragement each comprise seven items; the Teaching domain consists of eight items.

Evidence of the PICCOLO's validity is based on data derived from two research samples—the EHSREP and the Bilingual Early Language and Literacy Supports (BELLS) project. Collectively, the PICCOLO was validated on more than 4,500 videotaped interactions of 2,199 parents and their 10- to 40-month-old children obtained across repeated assessments. Although not intended to be nationally representative, the study samples represent ethnically diverse, low-income families, with 41 percent of parents identified as European American; 37 percent as African American, and 22 percent as Latino. Approximately 60 percent of mothers were teens at the time of the focus child's birth, and up to two-thirds of mothers had at least a high school diploma or equivalent (52 percent for African Americans, 66 percent for European Americans, 28 percent for Latinos).

Psychometric evidence supports the reliability and validity of the measure (Roggman et al. 2009). Average interobserver agreement is 80 percent for Affection, 76 percent for Responsiveness, 83 percent for Encouragement, and 69 percent for Teaching. Within each domain, factor loadings are in the moderate to high range, and internal consistency reliability coefficients range from 0.75 to 0.80. Intercorrelations among the domains are moderate to strong in magnitude (0.40 to 0.73), and cross-age associations within domains demonstrate moderate stability over time (0.36 to 0.52). There is also evidence of the PICCOLO's convergent and predictive validity. Specifically, the domains of parenting measured by the PICCOLO related significantly to measures of supportive parenting derived from the Early Head Start Parent-Child Interaction Rating Scales for the Three-Bag Assessment (ACF 2002), including parental sensitivity (0.31 to 0.50), cognitive stimulation (0.26 to 0.56), and positive regard (0.31 to 0.57). Each domain also related to children's outcomes both concurrently and over time, including measures of children's cognitive development, emotion regulation, vocabulary production, receptive language, emergent literacy, and problem solving (Roggman et al. 2009).

The final coding team included five members of the original team of coders who were trained by the certified trainer. Inter-rater reliability between the team leaders and the five-member coding team was established on the 29-item binary scale to a criterion of 80 percent exact agreement. We conducted weekly inter-rater reliability checks throughout the coding period on three to five randomly selected videos. A total of 53 videos (10.5 percent of the 505 codable videos) served as

reliability videos.[24] On the binary scale, agreement across all coders averaged 85 percent, with a range of 81 to 87 percent (see Appendix B for details on the binary coding scheme). Agreement across observations for the domain scores were 91 percent for Affection; 80 percent for Encouragement; 80 percent for Responsiveness; and 87 percent for Teaching. Inter-rater agreement on the three-point scale yielded lower estimates than were obtained on the binary scale. Across all coders, agreement averaged 77 percent overall, with a range of 73 to 80 percent. Agreement averaged 87 percent for Affection, 72 percent for Encouragement, 72 percent for Responsiveness, and 76 percent for Teaching.

Based on recommended practices for scoring in research-based settings (personal communication with Roggman, April 2, 2010), scores for each item were collapsed into the binary scale described above. Within each of the four domains, overall scores (ranging from 0 to 1) were calculated by taking the mean of all component items. Kuder-Richardson (KR20) alpha coefficients were 0.62 for Affection, 0.51 for Encouragement, 0.65 for Responsiveness, and 0.47 for Teaching. Associations among the four domains were moderate in magnitude, ranging from 0.36 to 0.56 (all significant). An overall score of positive parenting was derived by calculating a mean score across each of the four domain scores (Cronbach's alpha = 0.77). To allow for comparisons to the EHSREP sample, domain and total scores were also computed by averaging scores on the original three-point scale, although reliability estimates were somewhat lower than those obtained using the binary scale. The authors also provide age-based scoring rubrics which specify cut-points for low, moderate, and high levels of support for each of the four PICCOLO domain scores (Roggman et al. 2009). The high range is represented by total scores greater than or equal to 11 (out of 14) on Affection and Responsiveness; scores greater than or equal to 10 (out of 14) on Encouragement; and scores greater than or equal to 8 (out of 16) on Teaching.

Measures of Parent Mental Health

Center for Epidemiologic Studies Depression Scale (CES-D). The CES-D is a self-administered screening tool used to identify symptoms of depression or psychological distress (Radloff 1977). The full version of the CES-D consists of 20 items, and the short form (CESD-SF) (Ross et al. 1983) consists of 12 items. Respondents are asked to rate how often each of the items applied to them in the past week on a four-point scale from "rarely or never" (score of 0) to "most or all of the time" (score of 3). Symptoms include poor appetite, restless sleep, loneliness, sadness, and lack of energy. Raw scores range from 0 to 36 for the short form, with higher scores indicating more depressive symptoms.

The CESD-SF has been used as a measure of parent well-being in large-scale studies such as the EHSREP and the Head Start Family and Child Experiences Survey (FACES). We chose the CESD-SF because of its use in previous Early Head Start studies, well-established psychometric properties, and short administration time.

Parents with scores on the CESD-SF of 15 or higher are considered as having severe depressive symptoms; those with scores of 10 to 15 are considered as having moderate depressive symptoms; and those who score between 5 and 10 are considered as having mild depressive symptoms.

---

[24] We received a total of 522 videos of Two-Bag administrations from the field. We were unable to code 13 using the PICCOLO due to a number of reasons, including the parent speaking a language other than English or Spanish (*n* = 7), poor audio quality (*n* = 3), poor lighting (*n* = 1), the parent and/or child being out of frame (*n* = 1), and the interaction not being recorded due to field interview error (*n* = 1). An additional 4 videos were shorter than the required duration and thus excluded from the analyses. We were therefore able to compute scores from a total of 505 videos.

Table C.4 illustrates the average, standard deviation, range, and internal consistency of CESD-SF scores among parents in the Baby FACES study.

The Parenting Stress Index—Short Form (PSI-SF). The PSI-SF measures the degree of stress in the parent-child relationship stemming from three sources: (1) the child's challenging temperament, (2) parental depression, and (3) negative reinforcement of parent-child interactions (Abidin 1995). We employed the PSI-SF due to its previous use in the EHSREP and ease of administration. We included the Parental Distress and Parent-Child Dysfunctional Interaction subscales in Baby FACES.

The Parenting Stress Index—Short Form (PSI-SF). The PSI-SF measures the degree of stress in the parent-child relationship stemming from three sources: (1) the child's challenging temperament, (2) parental depression, and (3) negative reinforcement of parent-child interactions (Abidin 1995). We employed the PSI-SF due to its previous use in the EHSREP and ease of administration. We included the Parental Distress and Parent-Child Dysfunctional Interaction subscales in Baby FACES.

The Parental Distress subscale (five items) measures the level of distress the mother or father is feeling in his or her role as a parent, including a low sense of competence and a high level of stress due to perceived restrictions stemming from parenting. The parent answers whether or not he or she agrees with statements such as "You have been unable to do new and different things," and "You feel trapped by your responsibilities as a parent." Parents rate each item on a five-point scale from "strongly disagree" to "strongly agree." Scores can range from 5 to 25. Higher scores indicate higher levels of parental distress.

The Parent-Child Dysfunctional Interaction subscale (six items) measures a parent's perception that his or her child does not meet expectations and that interactions with the child are not reinforcing to the parent. The parent answers whether he or she agrees with statements such as "Most times, you feel that your child does not like you and does not want to be close to you" and "When you do things for your child, you get the feeling that your efforts are not appreciated very much." Parents rate each item on a five-point scale from "strongly disagree" to "strongly agree." Scores can range from 6 to 30. Higher scores indicate a more dysfunctional parent-child interaction.

Table C.4 illustrates the average, standard deviation, range, and internal consistency of PSI scores among parents in the Baby FACES study.

Table C.4 Parent Mental Health When Focus Child is 2 Years Old

Outcome	Possible Range		Reported Range		Mean/ Percentage	Standard Deviation	Cronbach Alpha
	Min.	Max.	Min.	Max.			
CESD-SF Raw Score	0	36	0	36	3.92	5.74	0.92
CESD-SF: Severe Depressive Symptoms	0	1	0	1	5.23	22.28	n.a
CESD-SF: Moderate Depressive Symptoms	0	1	0	1	7.23	26.73	n.a
CESD-SF: Mild Depressive Symptoms	0	1	0	1	15.45	36.19	n.a
CESD-SF: No Depressive Symptoms	0	1	0	1	71.59	45.15	n.a
PSI: Parental Distress	5	25	5	25	10.48	4.79	0.81

Outcome	Possible Range		Reported Range		Mean/ Percentage	Standard Deviation	Cronbach Alpha
	Min.	Max.	Min.	Max.			
PSI: Parent-Child Dysfunctional Interaction	6	30	6	30	8.68	4.65	0.84
Sample Size							
Parent Interview			**440-521**				

Source: Spring 2010 Parent Interview.

Note: Sample restricted to the 1-year-old Cohort. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 2.

n.a.=not applicable; PSI = Parenting Stress Index; CESD-SF = Center for Epidemiologic Studies Depression Scale Short Form. Severe depressive symptoms = scores of 15 or higher; moderate depressive symptoms = scores of 10 or higher but lower than 15; mild depressive symptoms = scores of 5 or higher but lower than 10; no depressive symptoms = scores lower than 5.

Measure of Functioning of Families

The Family Environment Scale, Family Conflict Subscale (FES). The FES measures the extent to which the open expression of anger and aggression and conflict-filled interactions are characteristic of the family (Moos 2002). Parents rated five items on a four-point scale, where a 4 indicates higher levels of agreement with statements such as "We fight a lot" and "We sometimes hit each other." Scores can range from 1 to 4. The subscale score is then the mean of the five individual item scores. We included the FES because it had been previously included in the EHSREP. For the Baby FACES sample, however, we removed one item: "We hardly ever lose our tempers." It had a low correlation with the rest of the items in the scale and therefore reduced the overall alpha of the measure.

Table C.5 illustrates the average, standard deviation, range, and internal consistency of FES scores among parents in the Baby FACES study.

Table C.5 Functioning of Families at Age 2

Outcome	Possible Range		Reported Range		Mean/ Percentage	Standard Deviation	Cronbach Alpha
	Min.	Max.	Min.	Max.			
FES-Family Conflict[a]	1	4	1	3	1.49	0.43	0.56
Sample Size							
Parent Interview			**428**				

Source: Spring 2010 Parent Interview.

Note: Sample restricted to the 1-year-old Cohort. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 2.

FES = Family Environment Scale.

Measures of Home and Neighborhood Environment

Home Observation for Measurement of the Environment. The HOME measures the quality of stimulation and support available to a child in the home environment (Caldwell and Bradley 1984). It has separate inventories for infants and toddlers (birth to 3 years old), early childhood (ages 3 to 6), and middle childhood (ages 6 to 10). Information needed to score theinventory is obtained through a combination of interview and 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 Project on Human Development in Chicago Neighborhoods (PHDCN). We derived five subscales from this assessment, as well as the total score:

- ***Emotional Responsitivity*** measures responsive and supportive parenting behavior observed by the interviewer during the home visit. Interviewer observations of the parent and child during the interview inform the items in this subscale, and explore such questions as whether the mother praised the child, whether she expressed warmth and affection toward the child, and whether she responded verbally to the child's verbalizations during the interview.

- ***Maternal Verbal-Social Skills*** measures the parent's ability to speak freely and clearly to the interviewer. This subscale comprises interviewer observations of the parent during the interview.

- ***Support of Cognitive, Language, and Literacy Environment*** measures the provision of a variety of developmentally stimulating toys and furnishings, as well as whether the parent provides toys for the child during the visit, reads to the child several times per week, and talks to the child while doing household chores. Items are obtained by a combination of parent report and interviewer observation. We also created another measure of **Enhanced Cognitive, Language, and Literacy Environment** by crediting parents only when more toys are provided to the child.

- ***Absence of Punitive Interactions*** measures harsh or punitive parenting behavior observed during the home interview. Interviewer observations of the parent and child during the interview inform the items in this subscale, and include such events as shouting at, expressing annoyance or hostility toward, hitting, scolding, or restricting the child. Items received a score of 1 if the parent did not engage in harsh or punitive behaviors during the home visit.

- ***Internal Physical Environment*** measures the cleanliness, organization, and warmth of the home environment. This subscale comprises interviewer observations during the interview.

- ***Total Score*** measures the cognitive stimulation and emotional support provided by the parent in the home environment. It includes all 30 items used in the five previous subscales.

We also derived an **External Environment score** (not included in the HOME total), which measures the physical and social environment of the face block[25] where the family lives based on some neighborhood rating items from the PHDCN. This subscale comprises interviewer

---

[25] A face block is defined as the two sides of one street between intersecting streets.

observations of the neighborhood. Examples include the general condition of most of the housing units, presence of 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.

The internal consistency reliability coefficients (Cronbach's alphas) were 0.84 for the original infants and toddlers HOME inventory and ranged from 0.49 to 0.78 for the subscales. Kuder-Richardson coefficients were 0.89 for the inventory and ranged from 0.44 to 0.89 for the subscales. The test-retest reliability estimates were 0.77 for the inventory and ranged from 0.30 to 0.77 when administered at ages 12 and 24 months. The intraclass correlation, which measures stability by comparing the similarity of paired scores relative to the total variation of all scores, resulted in slightly lower values. The intraclass correlation coefficients were 0.76 for the inventory and 0.30 to 0.76 at ages 12 and 24 months. The inter-rater reliability estimates ranged from 0.76 to 1.0 for the HOME.

Families' HOME inventory scores administered when the child was 6, 12, and 24 months old were compared with the child's scores on the Bayley Scales of Infant Development Mental Development Index (MDI) at 6 months and 12 months, the Stanford-Binet at 36 months and 54 months, and the Illinois Test of Psycholinguistic Abilities (ITPA) at 37 months. The HOME was found to be a better predictor of intelligence than socio-economic measures and was a stronger predictor for females and whites. The HOME was also compared with the Supplement to the HOME for Impoverished Families (SHIF), the Nursing Child Assessment Feeding Scale (NCAFS) and the Nursing Child Assessment Teaching Scale (NCATS). Below are the comparisons between the measures:

- Comparison with the Bayley MDI: The correlations between the HOME inventory score at 6 months and the Bayley MDI at 6 months and 12 months were 0.14 and 0.16 (subscale correlations ranged from 0.01 to 0.27), respectively. The correlation between the HOME at 12 months and the Bayley MDI score at 12 months was 0.30 (subscales ranged from 0.01 to 0.28).

- Comparison with the Stanford-Binet: The correlations between the HOME inventory score at 6 months and the Stanford-Binet at 36 months and 54 months were 0.50 (subscales ranged from 0.24 to 0.41) and 0.44 (subscales ranged from 0.10 to 0.44), respectively. The correlation between the HOME at 12 months and the Stanford-Binet at 36 months was 0.58 (subscales ranged from 0.24 to 0.56), respectively. The correlations between the HOME at 24 months and the Stanford-Binet at 36 months and 54 months were 0.71 (subscales ranged from 0.41 to 0.64) and 0.57 (subscales ranged from 0.28 to 0.56), respectively.

- Comparison with the ITPA: The correlations between the HOME inventory scores at 6 months and 24 months and the total ITPA score at 37 months were 0.39 and 0.61, respectively.

- Comparison with SHIF: The correlation between the HOME and the SHIF was 0.69.

- Comparison with the NCAFS and the NCATS: In a nonrepresentative sample of impoverished urban families, the correlations were 0.55 and 0.42 between the HOME and the NCAFS and NCATS, respectively.

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 variables as the general condition of most of the housing units, garbage in the street or on the sidewalk, traffic volume, and people arguing or fighting in the street. The scale score is the mean of the item z-scores. Higher scores indicate more disorder.

Table C.6 illustrates the average, standard deviation, range, and internal consistency for the HOME and neighborhood disorder scores of families with 2-year-olds in the Baby FACES study.

The Confusion, Hubbub, and Order Scale (CHAOS). The CHAOS scale assesses the level of confusion and disorganization in the child's home environment (Matheny et al. 1995). Parents completed the CHAOS scale in the self-administered questionnaire of the parent interview. The scale consists of 15 statements, to each of which a parent or caregiver responds on a four-point scale ranging from 1 (very much like your own home) to 4 (not at all like your own home). A single scale score is derived from the CHAOS scale by summing the responses for the 15 items. Scores for each response can range from 0 to 45. A higher score indicates a more chaotic, disorganized, and hurried home. The unweighted mean of the CHAOS score is 11.23 (SD = 6.26) for the homes of 2-year-olds in the Baby FACES sample.

The internal consistency (Cronbach's alpha) for the entire scale was 0.79. The test-retest reliability (12-month interval) for the total test score was 0.74. To examine concurrent validity, the authors of the CHAOS scale compared their scale with the physical and social environment codes in the Purdue Home Simulation Inventory (PHSI). (PHSI codes are completed by trained observers.) The authors reported that the correlations between the CHAOS scale and several of the PHSI social environment codes were in the moderate range: physical interference ($r = -0.36$), number of known objects named ($r = -0.38$), and ignoring the child's bids for attention ($r = 0.45$). Together, the PHSI social environment codes explained 59 percent of the variance in the CHAOS scores. The correlations between the CHAOS scale and several of the PHSI physical environment codes also were in the moderate range: number of siblings ($r = 0.55$) and number of rooms per person ($r = -0.33$). Together, the PHSI physical environment codes explained 39 percent of the variance in the CHAOS scores.

Table C.6 illustrates the average, standard deviation, range, and internal consistency for the CHAOS scores of families with 2-year-olds in the Baby FACES study.

Table C.6 Home and Neighborhood Environment at Age 2

Outcome	Possible Range Min.	Possible Range Max.	Reported Range Min.	Reported Range Max.	Mean/ Percentage	Standard Deviation	Cronbach Alpha
HOME							
Parental Warmth	0	7	0	7	6.09	1.36	0.73
Verbal/Social Skills	0	3	0	3	2.84	0.52	0.61
Parental Lack of Hostility	0	5	0	5	3.67	2.00	0.95
Support of Cognitive, Language, and Literacy Environment	0	12	4	12	9.61	1.25	0.48
Enhanced Cognitive, Language, and Literacy Environment	0	12	1	12	8.03	2.19	0.64
Internal physical environment	0	3	0	3	2.29	0.80	0.31
HOME Total Score	0	30	13	29	24.55	3.22	0.69

Outcome	Possible Range		Reported Range		Mean/ Percentage	Standard Deviation	Cronbach Alpha
	Min.	Max.	Min.	Max.			
CHAOS Total Score	0	45	0	31	11.23	6.26	0.79
Neighborhood Disorder	n.a	n.a.	-1	2	0.00	0.65	0.80
Sample Size							
Parent Interview			**471-540**				

Source: Spring 2010 Parent Interview.

Note: Sample restricted to the 1-year-old Cohort. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 2.

Exposure to Violence. The Exposure to Violence scale measures how many violent incidents (out of four) a child has observed in his or her lifetime. Items come from the ITSEA (Carter and Briggs-Gowan 2000), which asks parents to respond yes or no to questions that explore, for example, whether a child has "seen violence in their neighborhood" or "seen someone hit, push or kick a family member."

Measures of Home Visit and Classroom Quality

Home Visit Rating Scale-Adapted. The HOVRS-A[26] assesses a variety of dimensions of home-visiting quality and content, including home visitor responsiveness, nonintrusiveness, support of parent-child interaction, and parent and child engagement in the visit (Roggman et al. 2009). The HOVRS-A consists of seven items that are rated from 1 to 5, with anchors of 1 (minimal), 3 (moderate), and 5 (good practice). The scores can be combined to form a total score and two subscale scores:

- **Visitor Strategies Quality** focuses on the home visitor's responsiveness to the parent and child and consists of four items: (1) home visitor facilitation of parent-child interaction; (2) home visitor-family relationship; (3) home visitor responsiveness to family; and (4) home visitor nonintrusiveness.

- **Visitor Effectiveness Quality** assesses the parent's and child's engagement with each other and with the home visitor and consists of three items: (1) parent-child interaction during home visit; (2) parent engagement during home visit; and (3) child (infant or toddler) engagement during home visit.

---

[26] Four main modifications were made in creating the HOVRS-A from the HOVRS. First, to make the measure easier to score, the number of scale rating points was reduced from seven to five. This step helped to establish inter-rater reliability, because there are fewer subtle distinctions to make between one rating point and another. Second, the indicators were aligned across each of the three anchors (1, 3, and 5) to ensure that they are consistent and that the same types of behaviors are assessed at each level. Third, the Home Visitor Relationship with Family item was adapted so that it taps both the home visitor's engagement and relationship with the family and the family's relationship with the home visitor. Finally, we created two versions of the last item, Child Engagement During Home Visit, one for visits with a focus child up to 12 months old (Infant Engagement During Home Visit) and another for visits with toddlers 12 to 24 months (Toddler Engagement During Home Visit), and ensured that the indicator wording on all items is appropriate for infants and toddlers.

The overall quality score has high internal consistency (0.82). However, the Effectiveness Quality subscale has somewhat lower internal consistency (0.68). This score is slightly lower than the 0.70 standard in the field but higher than estimates reported by Peterson and Roggman (2006).

Table C.7 illustrates the average, standard deviation, range, and internal consistency for the HOVRS-A scores for home visits observed during the spring 2010 Baby FACES data collection.

Table C.7. Summary Statistics for Baby FACES Child Care Quality Data: HOVRS-A

| | Possible Range | | Reported Range | | | | |
HOVRS-A Scales	Min.	Max.	Min.	Max	Mean	Standard Deviation	Cronbach's Alpha
HOVRS-A Overall Quality	1	5	1.57	4.71	3.40	0.77	0.82
Visitor Strategies Quality	1	5	1.25	5.00	3.26	0.81	0.76
Effectiveness Quality	1	5	1.00	5.00	3.58	0.89	0.68
Sample Size	**189**						

Source: Spring 2010 Home Visit Observations.

Note: Includes observations of home visits to the Newborn and 1-year-old Cohort in spring 2010.

HOVRS-A = Home Visit Rating Scale-Adapted.

Infant/Toddler Environment Rating Scale-Revised (ITERS-R). The full ITERS-R (Harms et al. 2003) consists of 39 items organized under seven subscales: (1) Space and Furnishings (5 items), (2) Personal Care Routines (6 items), (3) Listening and Talking (3 items), (4) Activities (10 items), (5) Interaction (4 items), (6) Program Structure (4 items), and (7) Parents and Staff (7 items). Items on ITERS-R are rated from 1 to 7, with the authors providing rating descriptors of 1 (inadequate), 3 (minimal), 5 (good), and 7 (excellent). For observations of classrooms serving 1-year-old children (those in the Newborn Cohort), the Baby FACES study used a modified 32-item ITERS-R scale that excluded all Parents and Staff subscale items (which rely heavily on staff reports rather than observation). Classroom observations also included counts of children and the adults caring for them that we used to compute child-adult ratios and group sizes. We computed each classroom's total ITERS-R score by averaging the scores on all items collected for that classroom, and computed the six mean subscale scores for each classroom by averaging the classroom's scores on the items in each subscale.

Table C.8 illustrates the average, standard deviation, range, and internal consistency for the ITERS-R scores for home visits observed during the spring 2010 Baby FACES data collection.

Table C.8 Summary Statistics for Baby FACES Child Care Quality Data: ITERS-R

| | Possible Range | | Reported Range | | | | |
ITERS-R Scales	Min.	Max.	Min.	Max	Mean	Standard Deviation	Cronbach's Alpha
ITERS-R Total	1	7	1.97	5.29	3.69	0.74	0.85
Personal care	1	7	1.00	6.50	3.06	1.22	0.73
Furnishings	1	7	1.60	5.80	3.74	1.01	0.53
Listening and talking	1	7	2.00	5.67	3.86	0.89	0.28
Activities	1	7	1.75	4.90	3.47	0.78	0.55
Interaction/social	1	7	1.75	6.50	4.47	1.20	0.79

ITERS-R Scales	Possible Range		Reported Range			Standard Deviation	Cronbach's Alpha
	Min.	Max.	Min.	Max	Mean		
Program structure	1	7	1.00	7.00	4.17	1.41	0.59
Sample Size	**43-53**						

Source: Spring 2010 Classroom Observations.

Note: Includes observations of classrooms to the Newborn Cohort in spring 2010.

ITERS-R = Infant/Toddler Environment Rating Scale-Revised.

Classroom Assessment Scoring System-Toddler. The CLASS-T is an adaptation of the Pre-K CLASS (Pianta et al. 2008) that focuses on teacher-child interaction quality in toddler child care classrooms (Pianta et al. 2010a). Compared with other established quality measures of global or structural quality, the CLASS-T measures process quality along eight dimensions: (1) Positive Climate, (2) Negative Climate, (3) Teacher Sensitivity, (4) Regard for Child Perspectives, (5) Behavior Guidance, (6) Facilitation of Learning and Development, (7) Quality of Feedback, and (8) Language Modeling, within two broad 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), middle (3-5), and high (6-7) range.

We used the CLASS-T for classroom observations of 1-year-old Cohort children at age 2. Observations also included counts of infants and toddlers and the adults caring for them that we used to compute child-adult ratios and group sizes. We computed dimensions scores by averaging ratings obtained across four independent observation cycles, and calculated domain scores for each classroom by averaging the scores for the component dimensions on which the domain scores were based. (See Appendix D for results of factor analyses.) We reverse-coded scores on Negative Climate prior to calculating the domain mean score.

To ensure the reliability of our data, we assessed the inter-rater reliability of the field staff who observed classrooms serving 2-year-old children using the CLASS-T. Overall, reliability estimates were in accord with standards established by the measure developers—80 percent agreement (the same score or within one point) with the scores of the developer-certified gold standard group leaders. As Table C.9 shows, trained observers scored within 1 rating point of gold standard observers' scores, who rated the same classrooms on each of the seven-point CLASS-T dimensions, 91 to 100 percent of the time. Observers were thus adequately trained and maintained an adequate degree of reliability in CLASS-T scoring throughout Baby FACES data collection.

Table C.9. Gold Standard Reviewers and Field Staff Show: Close Agreement on CLASS-T Dimension Scores (Percentages)

CLASS-T Dimension Scores	Agreement Within 1 Point
Positive Climate	0.93
Negative Climate	1.00
Teacher Sensitivity	0.91
Regard for Child Perspectives	0.93
Behavior Guidance	0.95
Facilitation of Learning and Development	0.92
Quality of Feedback	0.95
Language Modeling	0.95
Sample Size	**14**

Source: Spring 2010 classroom observations.

As Table C.10 shows, the internal consistency reliability of the CLASS-T Emotional and Behavioral Support and Engaged Support for Learning domain scores in our field observation was 0.89 and 0.94, respectively. This is comparable to estimates reported for the domain of Emotional Climate (0.88) in a pilot study of the adapted measure conducted in 30 toddler classrooms (Thomason and LaParo 2009).[27] As reported by the study authors, evidence of the validity of the adapted measure was supported by associations between a number of CLASS-T dimension scores and other indicators of quality. Specifically, observed scores on Positive Climate and Teacher Sensitivity were most consistently associated with characteristics of the classroom, including teacher education level, group size, and child-teacher ratios (correlations range from 0.33 to -0.61).

Table C.10. Summary Statistics for Baby FACES Child Care Quality Data: CLASS-T at Age 2

CLASS-T Scales	Possible Range		Reported Range		Mean	Standard Deviation	Cronbach's Alpha
	Min.	Max.	Min.	Max			
Emotional and Behavioral Support	1	7	2.80	6.90	5.28	0.83	0.89
Positive Climate	1	7	2.25	7.00	5.50	1.17	0.94
Negative Climate	1	7	1.00	2.75	1.27	0.38	0.65
Teacher Sensitivity	1	7	1.75	7.00	4.81	1.05	0.89
Regard for Child Perspectives	1	7	1.75	7.00	4.69	1.03	0.88
Behavior Guidance	1	7	1.50	7.00	4.68	1.15	0.92
Engaged Support for Learning	1	7	1.00	6.58	3.54	1.21	0.94
Facilitation of Learning and Development	1	7	1.00	7.00	3.89	1.14	0.85
Quality of Feedback	1	7	1.00	6.25	3.43	1.35	0.92
Language Modeling	1	7	1.00	6.50	3.32	1.33	0.90
Sample Size	**206-220**						

Source: Spring 2010 Classroom Observations.

Note: Includes observations of classrooms to the 1-year-old Cohort in spring 2010.

CLASS-T = Classroom Assessment Scoring System-Toddler

Table C.10 illustrates the average, standard deviation, range, and internal consistency for CLASS-T scores for classrooms observed during the spring 2010 Baby FACES data collection.

Measure of Quality of the Parent-Caregiver Relationship

Parent-Caregiver Relationship Scale. The PCRS assesses the perceived quality of the relationship between parents and the child's home visitor or teacher (Elicker et al. 1997). Parents reported on the quality of their relationship with the home visitor or teacher; staff, in turn, provided similar reports on their relationship with the parent. The PCRS is intended to provide focused information on multiple dimensions and specific perceptions of the dyadic relationship.

Items on PCRS focus on important dimensions of the parent-caregiver relationship, including trust and confidence, communication, respect/acceptance, caring, competence/knowledge, partnership/collaboration, and shared values. Respondents complete the questionnaire in reference to a specific caregiver or parent, indicating on a five-point scale their level of agreement or

---

[27] The measure used by Thomason and LaParo (2009) was a downward extension of the CLASS (Pianta et al. 2008) that included only six of the eight component dimensions. 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 between the measures 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).

disagreement with a statement. Example statements include, "If there is a problem, my child's teacher or home visitor and I always talk about it soon," and "If there is a problem, this child's parent and I always talk about it soon." Scale scores in Baby FACES reports represent the average across a subset of these items. Separate scores are calculated for staff (using six items) and parents (using seven items).

The full PCRS scale includes 35 items, which were narrowed down to the 13 items used for spring 2010 data collection. We shortened the measure in response to a call after spring 2009 data collection expressing concern about the appropriateness of some items for staff and to reduce the burden on respondents. We selected a subset of items for use in spring 2010 data collection with acceptable internal consistency and reliability and that focused on areas of importance.

In a study of 217 parents and caregivers (Elicker et al. 1997), the PCRS was correlated with aspects of the infant care environment, including the amount of time the infant received care from the caregiver and caregiver work satisfaction. The authors reported internal consistency reliabilities of 0.93 for parents and 0.94 for caregivers on the measure. Correlations among the parent and caregiver scales were not significant, however, suggesting that parent-caregiver reports were incongruent. Typically, on such measures of perceived relationships, caregivers rate parents lower than parents rate caregivers, with caregivers' responses varying with demographic characteristics of parents (such as, age, education, income, and marital status). The internal consistency reliability of the PCRS total score in our spring 2010 field observation for parents in home-based and center-based settings, respectively, was 0.93 and 0.90. It ranged between 0.89 and 0.91 for caregiving staff in centers and home-based settings.

Table C.11 presents the mean, standard deviation, range, and internal consistency for the spring 2010 PCRS scores.

Table C.11. Staff-Parent Relationship Quality at Age 2

	Possible Range		Reported Range			Standard	Cronbach's
PCRS Scales	Min.	Max.	Min.	Max	Mean	Deviation	Alpha
PCRS Scores for Children Served in Centers							
Parent Report	1	5	1.57	5.00	4.55	0.57	0.90
Teacher Report	1	5	1.00	5.00	4.20	0.79	0.89
PCRS Scores for Children Receiving Services by Home Visits							
Parent Report	1	5	1.00	5.00	4.58	0.59	0.93
Home Visitor Report	1	5	1.00	5.00	4.38	0.69	0.91
Sample Size	**216-301**						

Sources: Spring 2010 Parent, Teacher, and Home Visitor Interviews.

Note: Includes interviews with parents, teachers, and home visitors of the 1-year-old Cohort in spring 2010.

PCRS = Parent-Caregiver Relationship Scale.

APPENDIX D. ANALYTICAL ISSUES

To fully support our findings and inform our analytical decisions, we further explored various analytical issues, many of which arose as a result of our data collection administration and findings from the baseline report. This appendix focuses on seven main analytic issues: (1) comparison of the cohorts at baseline and when children in both cohorts are age 1, (2) analysis of the Family Services Tracking (FST) data, (3) use of the Classroom Assessment Scoring System-Toddler (CLASS-T) scores, (4) factor analysis of the CLASS-T, (5) alternative definition of economic risk, (6) changes in the administration of the Ages & Stages Questionnaires, Third Edition (ASQ-3) and (7) analysis of families who leave the program early.

Comparisons Between Cohorts

To improve the precision of our estimates, we considered combining data from both cohorts at the same age. For example, in 2010, we collected data from children in the Newborn Cohort who were now one year old. We could combine these data with data collected in 2009 from children in the 1-year-old Cohort at age 1 to strengthen the description of 1-year-olds participating in Early Head Start. However, the two cohorts must not be so largely dissimilar that combining the data would either inflate or mask differences across any subgroup of interest. Below, we describe our approach to determining whether differences across cohorts exist and our findings of those comparisons. Appendix E contains tables displaying unweighted data for all 1-year-olds in our sample.

To evaluate differences between the two cohorts, we tested the significance of the means or proportions of the two cohorts across more than 100 demographic, risk, and child development characteristics. We evaluated these comparisons at two time periods: in 2009 (baseline data), and when children in both cohorts were one year old (2009 for the 1-year-old Cohort and 2010 for the Newborn Cohort).[28] Overall, we found only a few comparisons were significant at any level (see Table D.1). In the 2009 comparison, 26 characteristics were significant or trended to significance: 5 at the 0.01 level, 14 at the 0.05 level, and 7 at the 0.10 level. At age 1, potentially the most relevant comparison, 15 characteristics were significant or trended to significance: 7 at the 0.01 level, 4 at the 0.05 level, and 4 at the 0.10 level. We note that several of the significant comparisons are with items that are naturally related, such as the number of adults contributing to household income, actual income, and percentage of poverty. We also note that the different sizes of the two cohorts may distort some comparisons.

The pattern of differences presents a mixed picture; in financial terms, the 1-year-old Cohort is more advantaged than the Newborn Cohort. However, in terms of education and parenting stress, there are indications that the Newborn Cohort is better off. We will briefly summarize the significant differences and differences trending towards significance.

Looking at basic household characteristics, there are a few small differences that reach significance or trend towards significance. There was a trend to suggest that homes in the 1-year-old Cohort have a slightly larger number of adults at baseline, but the reverse is true when both groups are age 1. The Newborn Cohort has a larger average household size when both cohorts are age 1. There are also a few differences in race and age. There are more black fathers in the Newborn Cohort, and a trend to suggest more white mothers and children in the 1-year-old Cohort. The ages

---

[28] Not all time period comparisons are available for each measure due to differences in data collection between 2009 and 2010.

of the parents in each cohort are different initially but, importantly, are not different when comparing their ages when their children are one year old. The number of immigrant parents is higher for the 1-year-old Cohort initially, but not significant when children in both groups are one year old.

A few noteworthy differences exist among factors related to income, employment, education, and risk. Families in the 1-year-old Cohort have a higher average income when children are age 1, possibly because they also have more people contributing to household income. This situation occurs despite the similarities in parental age, household size, and household structure observed in the age 1 comparison. Similarly, a larger proportion of families in the 1-year-old Cohort have an income between 51 and 100 percent of the poverty level, whereas families in the Newborn Cohort are more likely to have incomes between 0 and 50 percent of the poverty level. Rates of employment are similar at age 1, but at baseline, mothers of newborns are less likely to have worked in the past year and more likely to be unemployed. By age 1, Newborn Cohort mothers were still more likely to be unemployed but trends suggest they were also more likely to be in job training. A few differences trending towards significance in fathers' education at baseline disappear by age 1.

Despite differences in income, only one factor of financial or food security was significant at age 1: Families in the 1-year-old Cohort were more likely to report having been evicted, but did so at very low frequency. At baseline, trends suggest the Newborn Cohort reported less food security; these differences were gone by age 1.

Not unexpectedly, higher rates of depression exist among mothers of newborns at baseline. However, these differences are gone when both groups are age 1. Mothers of 1-year-old Cohort children reported significantly more dysfunctional parent/child interactions when both groups were at age 1. Trends suggest mothers of newborns were more likely to report ever having a substance abuse problem at baseline, but these differences were gone when the children were one year old.

Lastly, we looked at measures of child development when the children were one year old. We found no significant differences in any measure–the Brief Infant Toddler Social Emotional Assessment (BITSEA), the Ages & Stages Questionnaire, Third Edition (ASQ-3), or the MacArthur-Bates Communicative Development Inventories (CDI).

Table D.1. Comparison of Cohorts at Baseline and Age 1

Characteristic	Means at Baseline (Spring 2009)			Means at Age 1		
	Newborn Cohort	1-Year-Old Cohort	Significance	Newborn Cohort	1-Year-Old Cohort	Significance
Household Characteristics						
Mean number of adults in household	1.74	1.85	+ [a]	2.06	1.85	*
Mean number of children in household	2.49	2.51	n.s.[a]	2.60	2.51	n.s.
Average household size	4.25	4.38	n.s.	4.66	4.38	+
Child lives with						
Two biological parents	0.47	0.49	n.s.	0.44	0.49	n.s.
Birth mother only	0.52	0.48	n.s.	0.55	0.48	n.s.
Birth father only	0.00	0.01	* [a]	0.01	0.01	n.s.[a]
No biological parents	0.00	0.02	** [a]	0.01	0.02	n.s.[a]
Single mother	0.53	0.46	n.s.	0.54	0.46	n.s.
Mother's average age in years	24.28	25.90	**	25.63	25.90	n.s.
Teenage mother	n.a.	0.53		0.55	0.53	n.s.
Father's average age in years	26.99	28.96	**	28.08	28.96	n.s.
Mother's race						
White	0.34	0.41	+	0.32	0.41	+
Black	0.24	0.18	n.s.[a]	0.24	0.18	n.s.
Hispanic	0.37	0.36	n.s.	0.38	0.36	n.s.
Native American	0.01	0.01	n.s.[a]	0.01	0.01	n.s.[a]
Biracial	0.03	0.03	n.s.	0.03	0.03	n.s.
Other	0.02	0.01	n.s.[a]	0.02	0.01	n.s.[a]
Father's race						
White	0.25	0.33	* [a]	0.26	0.33	n.s.
Black	0.34	0.22	** [a]	0.33	0.22	* [a]
Hispanic	0.35	0.39	n.s.	0.35	0.39	n.s.
Native American	0.01	0.02	n.s. [a]	0.02	0.02	n.s.[a]
Biracial	0.05	0.02	n.s. [a]	0.03	0.02	n.s.[a]
Other	0.01	0.01	n.s. [a]	0.01	0.01	n.s.[a]
Child's race						
White	0.24	0.33	*	0.25	0.33	+
Black	0.23	0.18	n.s.	0.24	0.18	n.s.
Hispanic	0.39	0.39	n.s. [a]	0.37	0.39	n.s.
Native American	0.01	0.01	n.s. [a]	0.02	0.01	n.s.[a]
Biracial	0.11	0.06	+ [a]	0.10	0.06	n.s.[a]
Other	0.02	0.01	n.s. [a]	0.02	0.01	n.s.[a]
Parent's immigration status						
Both parents born in the United States	0.72	0.69	n.s.	0.72	0.69	n.s.
One parent born outside the United States	0.13	0.10	n.s. [a]	0.12	0.10	n.s.
Both parents born outside the United States	0.14	0.21	* [a]	0.17	0.21	n.s.
Mother's average number of years in the United States, if born elsewhere	9.40	9.96	n.s.	11.59	9.96	n.s.
Father's average number of years in the United States, if born elsewhere	11.27	12.74	n.s.	11.68	12.74	n.s.[a]

Characteristic	Means at Baseline (Spring 2009)			Means at Age 1		
	Newborn Cohort	1-Year-Old Cohort	Significance	Newborn Cohort	1-Year-Old Cohort	Significance
Languages spoken in the home						
English only	n.a.	0.68		0.69	0.68	n.s.
Spanish (only or primarily)	n.a.	0.16		0.11	0.16	n.s.
English (primarily) and Spanish	n.a.	0.13		0.19	0.13	n.s.
English (primarily) and other languages	n.a.	0.02		0.01	0.02	n.s.[a]
Other language (only or primarily)	n.a.	0.01		0.00	0.01	** [a]
Household Income						
Average number of people contributing to household income	1.55	1.62	n.s.[a]	1.41	1.62	** [a]
Average household income	$19,969	$25,029	n.s.[a]	$17,844	$25,029	** [a]
Household income						
$0-$9,999	0.31	0.23	*	0.34	0.23	* [a]
$10,000-$17,499	0.30	0.25	n.s.	0.26	0.25	n.s.
$17,500-$24,999	0.17	0.21	n.s.	0.14	0.21	n.s.[a]
$25,000 or more	0.22	0.31	*	0.26	0.31	n.s.
Household income as a percentage of the poverty level						
0-50	0.38	0.28	*	0.47	0.28	** [a]
51-100	0.34	0.41	*	0.24	0.41	**
101-130	0.13	0.12	n.s.	0.09	0.12	n.s.
131 or higher	0.15	0.19	n.s.	0.20	0.19	n.s.
Household member receives public assistance	0.68	0.71	n.s.	0.78	0.71	n.s.
Employment						
Mother employed in last 12 months	0.45	0.56	*	0.47	0.56	n.s.
Mother currently working	n.a.	0.40		0.38	0.40	n.s.
Mother currently in job training	0.05	0.06	n.s.[a]	0.12	0.06	+ [a]
Mother is not employed, in school, or in training	0.40	0.30	*	0.40	0.30	*
Father employed in last 12 months	0.82	0.81	n.s.	0.76	0.81	n.s.
Father currently working	0.58	0.64	n.s.	0.56	0.64	n.s.
Father currently in job training	0.09	0.05	+ [a]	n.a.	0.05	n.a.
Education						
Mother currently taking classes	0.28	0.30	n.s.	0.35	0.30	n.s.
Mother's highest education completed						
Less than high school	0.43	0.41	n.s.	0.38	0.41	n.s.
High school diploma or equivalent	0.34	0.31	n.s.	0.38	0.31	n.s.
Some college or AA	0.20	0.23	n.s.	0.19	0.23	n.s.
BA or higher	0.03	0.04	n.s.	0.06	0.04	n.s.[a]
Mother has no high school credential	0.43	0.41	n.s.	0.38	0.41	n.s.
Father currently taking classes	0.21	0.13	* [a]	0.16	0.13	n.s.
Father's highest education completed						
Less than high school	0.49	0.45	n.s.	0.42	0.45	n.s.
High school diploma or equivalent	0.40	0.38	n.s.	0.43	0.38	n.s.
Some college or AA	0.10	0.13	n.s.[a]	0.11	0.13	n.s.

Characteristic	Means at Baseline (Spring 2009)			Means at Age 1		
	Newborn Cohort	1-Year-Old Cohort	Significance	Newborn Cohort	1-Year-Old Cohort	Significance
BA or higher	0.02	0.05	* [a]	0.03	0.05	n.s.[a]
Food Security Risk Factors						
Worried food would run out	0.40	0.38	n.s.	0.30	0.38	n.s.
Food didn't last and didn't have money to get more	0.36	0.29	+	0.26	0.29	n.s.
Could afford to eat balanced meals	0.24	0.25	n.s.	0.22	0.25	n.s.
Relied on only a few kinds of low-cost foods to feed children because of financial reasons	0.27	0.26	n.s.	0.23	0.26	n.s.
Couldn't feed children a balanced meal for financial reasons	0.18	0.18	n.s.	0.16	0.18	n.s.
Financial Security Risk Factors						
Could not pay the full amount of rent or mortgage	0.33	0.31	n.s.	0.29	0.31	n.s.
Was evicted from home or apartment	0.04	0.07	n.s.[a]	0.01	0.07	** [a]
Could not pay the full amount of gas, oil, or electricity bills	0.37	0.32	n.s.	0.34	0.32	n.s.
Had services turned off by the gas or electric company, or oil company would not deliver oil	0.06	0.09	n.s.[a]	0.09	0.09	n.s.
Had service disconnected by the telephone company for nonpayment	0.21	0.19	n.s.	0.22	0.19	n.s.
Parent Depression and Stress						
CESD-SF raw score	6.44	5.21	* [a]	4.78	5.21	n.s.
Not depressed	0.49	0.60	**	0.67	0.60	n.s.
Mildly depressed	0.28	0.24	n.s.	0.18	0.24	n.s.
Moderately depressed	0.11	0.09	n.s.	0.07	0.09	n.s.
Severely depressed	0.13	0.07	* [a]	0.07	0.07	n.s.
Smoking during pregnancy	0.15	0.11	n.s [a]	n.a.	n.a.	
Smoking inside the home	0.11	0.18	n.s.	0.18	0.18	n.s.
Currently smoking	0.20	0.21	n.s.	0.23	0.21	n.s.
PSI: parent–child dysfunctional interaction raw score	n.a.	8.79		7.82	8.79	** [a]
PSI: parental distress raw score	n.a.	10.86		10.28	10.86	n.s.
Psychological Risk						
Moderately or severely depressed (CESD-SF)	0.23	0.16	* [a]	0.15	0.16	n.s.
Substance abuse	0.13	0.08	+ [a]	0.00	0.08	--[b]
PSI subscales one standard deviation or more above the mean	n.a.	0.28		0.26	0.28	n.s.
Low risk	n.a.	0.60		0.64	0.60	n.s.
Medium risk	n.a.	0.31		0.30	0.31	n.s.
High risk	n.a.	0.10		0.06	0.10	n.s.
Parent Modernity Scale						
Traditional beliefs raw score	n.a.	19.78		19.89	19.79	n.s.
Progressive beliefs raw score	n.a.	20.07		20.36	20.07	n.s.
Child Development						
Parent-reported BITSEA raw score						
Problem domain	n.a.	10.57		13.48	10.57	n.s.[a]

Characteristic	Means at Baseline (Spring 2009)			Means at Age 1		
	Newborn Cohort	1-Year-Old Cohort	Significance	Newborn Cohort	1-Year-Old Cohort	Significance
Competence domain	n.a.	16.16		17.74	16.16	n.s.[a]
ASQ-3 raw score: total score	n.a.	216.23		219.01	216.23	n.s.
CDI (English) raw score						
Vocabulary comprehension	n.a.	30.26		29.51	30.26	n.s.
Vocabulary production	n.a.	2.74		2.13	2.74	n.s.[a]
CDI (Spanish) raw score[c]						
Vocabulary comprehension	n.a.	35.86		31.45	35.86	n.s.
Vocabulary production	n.a.	2.16		2.05	2.16	n.s.
CDI conceptual score (English and Spanish) [c]						
Vocabulary comprehension	n.a.	41.37		39.16	41.37	n.s.
Vocabulary production	n.a.	3.02		3.21	3.02	n.s.
Sample size	**139-174**	**572-683**		**92-130**	**572-683**	

Sources: Spring 2009 and 2010 Parent Interview.

Note: Comparisons at age 1 use baseline (2009) data from the 1-year-old Cohort and 2010 data from Newborn Cohort parents still enrolled in Early Head Start in spring 2010.

[a] Test of significance uses an assumption of unequal variances across cohorts.

[b] Questions on substance abuse were asked differently in 2009 and 2010. In 2009, parents were asked whether they ever had a substance abuse problem. In 2010, parents were asked whether they had a problem in the year since the 2009 survey. An age 1 comparison that uses 2009 data for the 1-year-old Cohort and 2010 data for the Newborn Cohort would not be appropriate.

[c] Spanish CDI scores reflect children from only bilingual households; sample size is 19 for the Newborn Cohort and 113 for the 1-year-old Cohort.

$^{+}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$

ASQ-3 = Ages & Stages Questionnaires, Third Edition; CDI = MacArthur-Bates Communicative Development Inventories; CESD-SF = Center for Epidemiologic Studies Depression Scale-Short Form; PSI = Parenting Stress Index.

n.a. = not available.

n.s. = not significant.

Family Services Tracking

The Family Services Tracking (FST) form was created to collect detailed information every week on the services children and families receive. By tracking these services weekly, we can examine in greater detail the nature and the intensity of the services that the average family in Early Head Start receives and how these services vary throughout the year.

FST data collection began in April 2009, a few weeks after the spring 2009 site visits for each program. At the onset of data collection, we gave each teacher and home visitor a list of children for whom they were responsible for reporting services, as well as a data manual and FAQ with instructions for filling out the forms. We offered teachers and home visitors two options for submission: (1) Use the secure FST website to fill out the forms weekly; or (2) complete a paper form and mail it to Mathematica's Princeton, New Jersey, office using a supplied business reply envelope. We worked with sites to further customize these options, when necessary. For example, some sites asked staff to fill out the hard copy forms and hand them to a central person to enter on the website. We permitted another site that used its own tracking information form to submit that form instead.

The FST data we analyze in this report pertains to services received by children in the 1-year old Cohort from the first week of July 2009 to the last week of June 2010. We defined a child as eligible for an FST report in each week that he or she was eligible to receive Early Head Start services. Therefore, a child who did not exit the Early Head Start program before June 2010 should have 52 FST reports.

Response rates varied by program, child, and time of year.

We employed several strategies to encourage timely and consistent submission of FST forms. During the first two weeks of July 2009, Mathematica coordinators made phone calls to sites to assess progress and address questions or concerns related to filling out the forms. After the first few months, coordinators sent a weekly email to teachers and home visitors to remind them to fill out the forms.

Despite efforts to maximize response rates, the number of records we received in each week varied throughout the year. Figure D.1 shows the number of eligible children and the number of reports received by week. Late program release dates, seasonal issues, and program exits explain much of the dynamics seen in the figure. We introduced the FST to programs on a rolling basis from April to September 2009, so a small number of programs with late release dates had not started submitting reports during the early weeks of the FST window.[29] We also observe a decline in the number of reports during the holiday season[30] and are missing weeks' worth of information for children who left Early Head Start before the end of June 2010.[31] Even after accounting for late program release dates, seasonal effects, and child eligibility, there are weeks for which we are missing data for reasons that were not immediately obvious.

---

[29] Children in the six programs that started using the FST after the first week of July 2009 were eligible for services but had missing reports.

[30] We observe a decline even though programs are supposed to submit reports when they are not operating.

[31] We assumed that a child was ineligible to receive services in the week he or she left the program unless a report was filed to indicate that he or she received services.
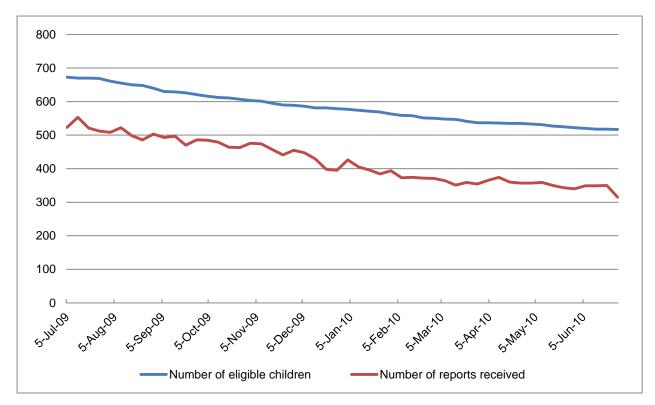
Figure D.1. Number of Eligible Children and Number of Reports Received, by Week

Variation in the number of reports per child is a concern, because missing data can bias our estimates of services received if the lack of information is systematically related to certain child or family characteristics. For example, if children with low attendance are more likely to be missing FST reports, estimates of average attendance will be biased upward, because the average does not include children with low attendance. Similarly, estimating the percentage of children or families who receive screenings, referrals, or attend parent education events in the period we observe is challenging, because a child may have received a screening in a week for which we are missing data. As a result, we may underestimate the percentage of children who received the service during the year. The following section describes the amount of missing FST data at the program- and child-levels.

Most programs submitted reports, but the number of reports we received for each child varied.

There was considerable variation in the extent to which programs filed reports for each sample child. We received at least one FST report from 87 of the 88 programs serving children in the 1-year-old Cohort;[32] the number of study children in each program ranges from one to 27. The percentage of children with at least one report submitted ranged from 0 to 100 percent. One program did not submit any reports during this period, and 41 programs submitted at least one report for all of their study children in the 1-year-old Cohort. Figure D.2 shows the distribution of the percentage of children with at least one report submitted across programs.

---

[32] One of the 89 programs in our sample serves only children in the Newborn Cohort.
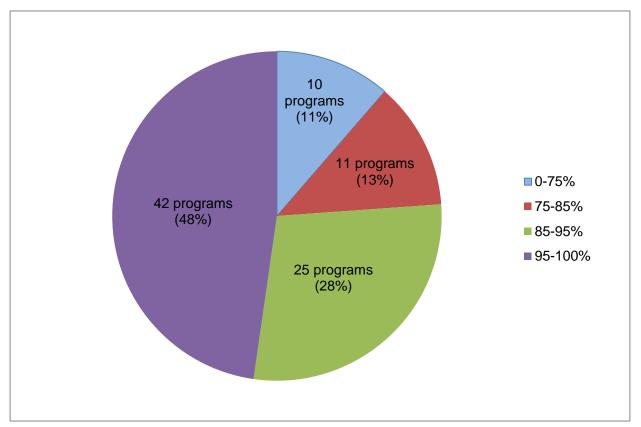
Figure D.2. Percentage of Enrolled Children Receiving at Least One FST Report, by Program

Within programs, there was also variation in the number of reports we received for each child. In July 2009, 779 children from the 1-year-old Cohort were eligible for an FST report.[33] Of this group, 673 children (86.4 percent) received at least one FST report. We used nonresponse weights to adjust for children with no reports submitted. Appendix A describes the construction of these weights.

In addition to the number of reports for each child, there was also variation in which weeks were missing and the length of the gaps between reports.[34] Among children with at least one FST report, we received reports for 72 percent of the weeks that a child was eligible during the 52-week period, on average. (Figure D.3 shows the distribution of this ratio for the 673 children with at least one submitted report.) The average number of gaps for children in the 1-year-old Cohort who had at least one FST report is 2.3. The number of gaps ranged from 0 to 12. Twenty-two percent of children had zero gaps. For children with at least one gap, 36 percent had a gap of no more than 5 weeks, and 32 percent of children had a gap of more than 9 weeks. The longest gap we observed for any child was 51 weeks. Table D.2 summarizes the data on the number and length of gaps.

---

[33] We include only children whose families consented to be in the study throughout the time period we observe.

[34] We define a gap as any period of missing data during the 52-week period.
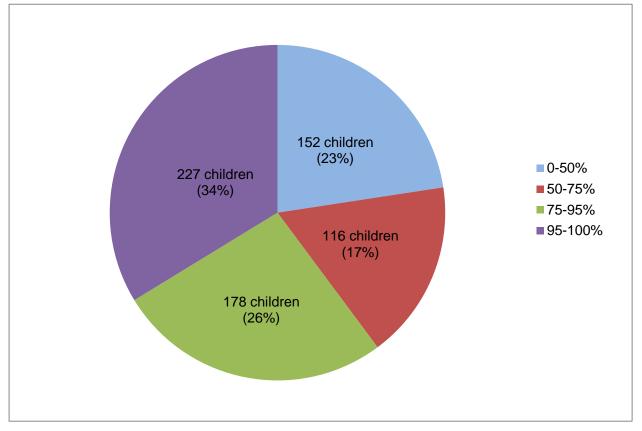
Figure D.3. Percentage of Eligible Weeks in Which an FST Report Was Received, by Child

Table D.2. Number and Length of Gaps in FST Reports

	Number of Children	Percentage
Zero Gaps	146	21.69
Gap of No Longer than Five Weeks	239	35.51
Gap of More than Five but No Longer than Nine Weeks	76	11.29
Gap of More than Nine Weeks	212	31.50
Total	**673**	100

Source: Family Service Tracking Reports, July 2009 to June 2010.

The frequency and length of gaps in FST reports could lead to bias in estimates of service receipt if the lack of information is nonrandom (that is, if it is related to attendance and service receipt, as well as to factors that influence attendance and service receipt). The following section investigates relationships between missing information and child, family, staff, and program characteristics.

Evidence that missing information is nonrandom.

Center attendance and home visit receipt are weakly correlated with the percentage of eligible weeks in which a child had a report (known as "percent reports"). But a stronger relationship exists between missing reports and the probability of receiving a screening or referral. Table D.3 shows correlations between the percent reports, with various indicators of service receipt. Although the correlation coefficients are relatively small in magnitude, the fact that most are significantly different from zero suggests that missing FST data is not a completely random event.

Table D.3. Pairwise Correlations Between Service Receipt and Percent Reports

	Correlation	*p*-value
Center Attendance Ratio[a]	0.02	0.73
Home Visit Ratio[a]	-0.12*	0.02
Received at Least One Referral	0.20**	0.00
Received at Least One Screening	0.26**	0.00
Attended at Least One Parent Education Event	0.06	0.12

Source: Family Service Tracking Reports, July 2009 to June 2010.

[a] Center attendance (home visit) ratio equals the number of center days attended (home visits completed) divided by the number of center days expected (home visits expected), ignoring missing values.

*p < 0.05; **p < 0.01

As Table D.4 shows, we also observe strong correlations between certain child and staff characteristics[35] and the amount of nonmissing data. The percentage of reports a child receives is significantly, or trending to significance, and positively correlated with being white and, having a disability diagnosis, and having a higher parent-reported staff-parent relationship score measured by the Parent-Caregiver Relationship Scale (PCRS). The percentage of reports received is significantly, or trending to significance and negatively correlated with being a dual language learner; having a mother who is not working, in school, or in training; and having a mother who had a baby as a teenager. In terms of staff characteristics, children whose providers report receiving more benefits and had more Early Head Start experience have a higher percentage of eligible reports. Children whose providers reported more depressive symptoms trend to have a lower percentage of eligible reports.

Meaningful correlations between observable characteristics and missing reports suggest that there may be other factors affecting report submission. These factors could also directly affect service receipt. Because of this possibility, we employed a multiple imputation strategy to minimize the bias in our estimates.

---

[35] We used child and staff characteristics measured in spring 2009, prior to the beginning of the FST period.

Table D.4. Correlations Between Percentage Reports and Child, Family, and Staff Characteristics

	Correlation	*p*-value
Service Stability	0.04	0.27
Number of service changes	0.03	0.38
Number of teacher changes	-0.29**	0.00
Number of gaps	-0.88**	0.00
Gap length	0.04	0.27
Child Characteristics		
Left Early Head Start program during FST window	-0.04	0.27
Age in months	0.04	0.27
Male	0.02	0.64
Dual language learner	-0.11**	0.00
White	0.11*	0.01
Any disability diagnosis	0.09*	0.04
Family Characteristics		
Mother not working, in school, or in training	-0.09*	0.03
Income-to-needs ratio	-0.03	0.54
Teen mother	-0.07+	0.08
Staff-Parent Relationship		
Parent-reported	0.08+	0.06
Staff-reported	0.05	0.17
Child Outcomes (Parent-Reported)		
ASQ, total score	0.01	0.85
BITSEA, problem domain	-0.03	0.48
BITSEA, competence domain	0.05	0.19
Child Outcomes (Staff-Reported)		
CDI, comprehension	0.01	0.84
CDI, production	0.01	0.81
BITSEA, problem domain	-0.13**	0.00
BITSEA, competence domain	0.02	0.63
Staff Characteristics		
CESD-SF, Categories	-0.08+	0.05
Number of employment benefits[a]	0.08*	0.03
Years of experience (total)	0.06	0.13
Years of experience (in Early Head Start)	0.09*	0.02
Bachelor's degree or more	0.05	0.24

Sources: Family Service Tracking Reports, July 2009 to June 2010; Parent Interview 2009; Staff-Child Report 2009; Staff Interview 2009.

[a] This variable contains the number of benefits that the staff member reported receiving: educational stipends for workshops, retirement/pension plan, life insurance, paid maternity leave, paid health insurance, dental insurance, paid sick leave, paid holidays, and paid vacations.

We used multiple imputation to account for missing data.

+*p* < 0.10; *p* < 0.05; **p* < 0.01

ASQ-3 = Ages & Stages Questionnaires, Third Edition; CDI = MacArthur-Bates Communicative Development Inventories; CESD-SF = Center for Epidemiologic Studies Depression Scale-Short Form.

We used multiple multivariate imputation to account for weeks in which a child was eligible for services but we did not receive a report. The imputation strategy we employed replaces missing values in the dataset by (1) regressing each variable with missing data on a set of covariates and (2) adding a stochastic component to predicted values such that the resulting variation in the imputed values mirrors the variation in the nonmissing data. We repeated this process multiple times, depending on the size and amount of missing data in each dataset. Box D.1 lists the variables included in our imputation model.

We imputed all time-invariant child, family, and staff characteristics separately before imputing FST data. We created 10 imputed datasets containing time-invariant characteristics and merged the raw FST dataset to each imputed dataset. We then conducted 2 imputations in each weekly FST dataset for a total of 20 imputations of the FST data.

We also included program fixed effects[36] to account for program-level differences and month fixed effects to account for seasonality. Program fixed effects allow us to predict the service use for a child with missing data using program characteristics common to children in the same program.

The estimates reported in Chapter IV are based on analyses conducted across the 20 multiply imputed datasets. We obtained each estimate by calculating the statistic of interest within each imputation and then averaging across imputations. The resulting standard errors account for two sources of variance: (1) sampling variance, which is based on average standard errors within each imputed dataset; and (2) imputation variance, which is the variance in estimated averages across imputations. Our estimates also adjust for differential probabilities of selection into the study sample.

Box D.1. Variables in Imputation Model

- ***Child demographic characteristics***
  - Race/ethnicity
  - Gender
  - Age in months
  - Dual language learner status
  - Number of child care arrangements

- ***Family demographic characteristics***
  - Number of children and adults in the home
  - Presence of biological parents in the home
  - Household income-to-needs ratio
  - Maternal employment status
  - Maternal education level
  - Teenage mother status
  - Child and family health insurance coverage

- ***Parental health***
  - Overall health status
  - Current and past drug, alcohol, and tobacco use
  - Receipt of treatment for a drug or emotional problem
  - Depressive symptoms (Center for Epidemiologic Studies Depression Scale [CES-D])
  - Parenting stress (Parenting Stress Index-Short Form [PSI-SF])

- ***Child health***
  - Child birth weight
  - Overall health status
  - Premature birth
  - Diagnosed disabilities

- ***Child outcomes***
  - MacArthur-Bates Communicative Development Inventories (CDI)—Infant Short Form
  - Brief Infant Toddler Social Emotional Assessment (BITSEA)
  - Ages & Stages Questionnaires, Third Edition (ASQ-3)

---

[36] We used program fixed effects when possible; when cell sizes were too small, we instead used program strata effects.

- **Staff characteristics**
  - Education level
  - Race/ethnicity
  - Years of experience teaching infants and toddlers
  - Years of experience in Early Head Start
  - Degree in early childhood education
  - Depressive symptoms
  - Quality of relationship with child's parents
  - Benefits received
  - Provider speaking a language other than English
- **Family program experiences**
  - Early Head Start service option in spring 2009
  - Number of times the family changed service options and providers during the observed FST period
  - Quality of relationship with Early Head Start provider
  - Typical number of center days attended or home visits received in the previous year
  - Whether the family left its Early Head Start program before June 2010
- **FST reporting quality**
  - Percentage of eligible weeks for which a child had a report submitted
  - Total number of FST reports
  - Number of gaps in FST reports
  - Largest gap observed
- **Weekly service variables**
  - Service option
  - Number of center days expected and attended (if in center-based or combination option)
  - Number of home visits expected and completed (if in home-based or combination option)
  - Whether the child attended a group socialization activity (if in home-based option)
  - Whether the child or family received a screening or referral
  - Whether parent attended a parent education event

CLASS-T Scores

We addressed a number of conceptual and analytic questions related to scoring the Classroom Assessment Scoring System-Toddler (CLASS-T; Pianta et al. 2010a). As noted in Appendix C, the CLASS-T is a downward extension of the CLASS (Pianta et al. 2008), which addresses teacher-child interaction quality in toddler child care classrooms. In Baby FACES, we used the CLASS-T in spring 2010 to observe interactions in classrooms serving 2-year-old children. As elaborated in Appendix C, there is preliminary evidence for the validity of the adapted measure, albeit in a relatively small sample of classrooms ($N = 30$; Thomason and LaParo 2009). Baby FACES represents the first large-scale effort to establish the reliability and validity of the measure, and by the end of the study, we will be able to document associations between the quality of teacher-child interactions and a broad range of child developmental outcomes.

Notably, since the spring 2010 Baby FACES data collection, the CLASS-T has undergone a number of modifications and refinements, resulting in a revised iteration of the instrument (Pianta et al. 2010b). Among the most significant changes to the instrument is the reconceptualization of the

domains that characterize teacher-child interaction quality. Specifically, the CLASS-T pilot manual (Pianta et al. 2010a) defines classroom interactions along eight dimensions grouped into three broader domains: Emotional Support (Positive and Negative Climate, Teacher Sensitivity, Regard for Child Perspectives), Classroom Organization (Behavior Guidance, Facilitation of Learning and Development), and Instructional Support (Quality of Feedback, Language Modeling). The authors note the validation of this "organization structure" using data from more than 3,000 classrooms ranging from preschool to fifth grade (Hamre et al. 2006). However, there exists less evidence for the use of this conceptualization in classrooms serving toddlers. While assessing teacher-child interactions along the same eight quality features, the revised CLASS-T manual (Pianta et al. 2010b) offers two overarching domains of toddlers' classroom experience: Emotional Support and Instructional Support. The Classroom Organization domain consists of a sole dimension in the revised model—Behavior Management; the dimension of Facilitation of Learning and Development, which was previously included as part of this domain, is now a component dimension of the Instructional Support domain. Other noteworthy changes to the CLASS-T include the provision of additional details, examples, and guidance to observers for the indicators within dimension ratings. During our spring 2010 data collection, the revised CLASS-T manual (Pianta et al. 2010b) was still under development and not yet available for use. Thus, the measure developers trained the field staff according to the guidelines, indicators, and exemplars in the pilot manual. We will continue to use this version of the instrument in our ongoing data collection.

Factor Analysis of CLASS-T Data

The modifications on the CLASS-T described above have noteworthy implications for the calculation of the domain scores; we thus conducted analyses to cross-validate the underlying factor structure of the CLASS-T in our sample and guide our approach to scoring at the domain level. Using data from 220 Early Head Start classrooms observed in spring 2010, we conducted a principal components factor analysis with varimax rotation using the eight component dimensions of the CLASS-T: (1) Positive Climate, (2) Negative Climate (reverse coded), (3) Teacher Sensitivity, (4) Regard for Child Perspectives, (5) Behavior Guidance, (6) Facilitation of Learning and Development, (7) Quality of Feedback, and (8) Language Modeling. We a priori retained two factors for rotation.[37] As Table D.5 shows, item loadings for the factors ranged from 0.71 to 0.94 and aligned closely with the domains identified by the CLASS-T developers (Pianta et al. 2010b). Of note, the dimension of Behavior Guidance loads onto the first factor; according to the author manual, this dimension is retained as an individual item that is not aggregated into either the Emotional Support or the Instructional Support domain scores. However, given the findings of the factor analysis, coupled with strong observed associations between Behavior Guidance and the dimensions corresponding to the Emotional Support domain, we retained Behavior Guidance as part of this first factor.[38] The two-factor solution demonstrated high internal consistency (alphas = 0.89 and 0.94 for the first and second factors, respectively) and explained a substantial portion of common variance. Thus, two composite scores, Emotional and Behavioral Support and Engaged

---

[37] In accord with the organizational structure of the broader CLASS-T domains in the revised manual, we also examined a three-factor solution. Behavior Guidance, however, did not emerge as a separate factor when a third factor was retained for rotation. Instead, Negative Climate loaded singly onto the additional factor, likely due to the low variability observed on this dimension (mean scores range from 1.00 to 2.75).

[38] The positive dimensions corresponding to the first factor were strongly and significantly correlated (0.63 to 0.86); associations between Negative Climate and the positive dimensions comprising this factor were moderate in magnitude (-0.30 to -0.45). Notably, associations between Behavior Guidance and the positive dimensions comprising the first factor ranged from 0.70 to 0.80. Associations among the three component dimensions corresponding to the second factor were strongly associated and statistically significant (0.80 to 0.91).

Support for Learning, were created by averaging the component items corresponding to each of the derived factors.[39]

Table D.5. CLASS-T Dimensions Load into a Two-Factor Solution, Spring 2010

	Factor	
Dimension	Factor 1: Emotional and Behavioral Support	Factor 2: Engaged Support For Learning
Positive Climate	0.74	0.30
Negative Climate	0.71	-0.07
Teacher Sensitivity	0.79	0.47
Regard for Child Perspectives	0.79	0.45
Behavior Guidance	0.82	0.37
Facilitation of Learning and Development	0.41	0.83
Quality of Feedback	0.18	0.94
Language Modeling	0.18	0.93
Mean (Standard Deviation)	5.30 (0.80)	3.54 (1.17)
Standardized Alpha[a]	0.89	0.94
Percentage of Total Variance Explained	63.15	15.95
Sample Size	**217**	

Source: Spring 2010 Baby FACES classroom observations.

Note: Includes observations of classrooms of the 1-year-old Cohort in spring 2010. Three of the 220 classrooms in the study were missing data for at least one dimension and were excluded from the factor analysis. As a result, the sample size was 217 classrooms.

[a] Standardized alpha calculated among items with loadings of 0.45 or higher.

CLASS-T = Classroom Assessment Scoring System-Toddler.

Alternatives for Measuring Economic Risk

In response to questions about the construction of the index measuring a family's economic risk, we explored three modifications to the measure relative to the index used in the Baby FACES 2009 report. The baseline Baby FACES economic risk index includes measures of risk for both financial and food security–the ability to pay bills and feed your family. We evaluated the effects of these modifications–one each to the individual scales and one to the construction of the index using both scales. First, we investigated whether adding a measure of poverty to the measure of financial risk would change the risk distribution among respondents. We also explored the possibility of reducing the number of categories included in the food security risk measure. Finally, we considered weighting unequally in the economic risk index the financial risk and food security risk measures. When necessary, we used 2009 baseline data for these analyses to allow for comparisons with the first report.

Including poverty indicator in financial risk. When exploring whether to include poverty in the measure of financial security, we found that it resulted in little change in the distribution of families across the scale measuring financial risk. The measure of financial security used in 2009 includes five factors:

---

[39] Based on the findings of this analysis and at the suggestion of the CLASS-T developers, we refer to the two derived composite domain scores throughout the report using these modified descriptors.

1. Whether you could pay your energy bills in full

2. Whether you could pay your rent or mortgage in full

3. Whether your phone had been disconnected

4. Whether your energy had been turned off

5. Whether you have been evicted

These items come from the Adult Well-Being Topical Module used by the Survey of Income and Program Participation (SIPP). In the baseline report, we scaled this measure in increments of two: families reporting no or one difficulty (low risk), two or three difficulties (medium risk), and four or five difficulties (high risk). In the 2009 baseline data, nearly 70 percent of families reported experiencing zero or one hardship. We explored whether adding a sixth factor–whether a family was living below the poverty line–would result in sizable changes in the distribution of the sample on this scale. We defined poverty as 100 percent of the poverty threshold, because most of our baseline sample (slightly more than two-thirds) fell into this category. When we add this indicator for poverty to the other five items, we see little change in the distribution of the sample across the scale; about 7 percent of the sample moves from low risk to medium risk and an additional 4 percent of families become high-risk.[40] In addition, correlations between the five questions and the poverty indicator are low (0 to 0.13). Measuring internal consistency of the six-item scale confirms that poverty status does not relate well to these items. Including poverty, the Cronbach's alpha for the scale is 0.63. But, if we eliminate the poverty indicator, the alpha rises to 0.67. Therefore, given the minimal changes that result and the weak relationship between the questions and a family's poverty status, we decided not to include the poverty indicator for 2010 to maintain the ability to compare financial risk over the duration of the study.

Reducing food security risk factors. We also considered the possibility of reducing the number of food security risk factors to one or two key issues (given the related nature of the items). The scale used in the 2009 report is drawn from an 18-item U.S. Department of Agriculture (USDA) module on food insecurity, which consists of three stages of questions. The first stage deals with perceptions and anxieties over food security; the second stage relates to reported instances of reduced food intake for adults; and the third stage covers reported instances of reduced food intake for children (Bickel 2000). The questions used in this study are the five questions of Stage One of the 18-item module:

1. Whether you were worried food would run out

2. Whether food did not last and you did not have enough money to buy more

3. Whether you could afford to eat balanced meals

4. Whether you could afford to feed the children balanced meals

5. Whether you relied on a few kinds of low-cost foods to feed the children, because the money was running out

The measure was scaled in increments of two: families reporting no or one difficulty (low risk), two or three difficulties (medium risk), and four or five difficulties (high risk). The five items are all highly correlated, ranging from 0.48 to 0.72. Measures of internal consistency reliability also show

---

[40] In this case, high risk includes the less than one percent of families reporting all six risk factors as well as those reporting four or five risk factors.

the items work together within the scale. Cronbach's alpha for the five-item scale is 0.88; removing any item would yield a lower alpha value. Lastly, exploratory and principal components factor analyses confirm a one-factor model. Given the relationships among the items, we kept the food security scale as it is currently measured, using all five items. This approach again allows for comparisons over time.

Weighting economic and food security for economic risk score. Finally, we considered whether the measures of financial and food security risk should be equally or unequally weighted in the ultimate index of economic risk, as it is possible that some individual risk factors are more severe than others. Based on the relationship among the questions included in each risk measure, we considered three options for creating the index, in some cases using more simplified scales for one or both measures. The economic risk index used in the Baby FACES 2009 report is an equally weighted measure of the number of affirmatives observed on both scales. Currently, a household is considered high-risk if the family answers yes to four or five questions on either the financial or food security scale; a household is considered medium-risk if the family answers yes to two or three questions on either scale; it is considered low-risk if the family answers yes to none of the questions or to one question on <u>both</u> scales. Based on the statistical relationship of the items within and between the two scales, we investigated the effects of collapsing one or both scales for the overall economic risk index.

For the financial security scale, we find a moderate relationship among the items that suggests one or two distinct concepts. Among all five items, the between-item correlations range from 0.16 to 0.53. Regarding the internal consistency reliability of the scale using all five items, we calculated an alpha of 0.67. We do, however, find that one item (eviction) does not relate as well to the other four items. This distinction is perhaps due to its rarity; only 6 percent of respondents reported being evicted in the last year. Removing eviction from the scale raises the value of the alpha to 0.70. Although the alpha including all five items is within the acceptable limits, we recognize that eviction is a rare but serious event. We therefore considered one scale option that groups eviction with the other four items and one option that treats it distinctly.

For the food security scale, as noted above, we find a very strong relationship between the five items. Removing any one factor would weaken the internal consistency of the scale. As also noted above, however, the five items comprise only the first stage of a three-stage module used by the USDA. In the full module, families progress from Stage One to Stage Two if they give an affirmative response to any of the first five items; those with no affirmative responses are considered to have a secure food supply (Bickel 2000). In the Baby FACES baseline data, about 55 percent of families reported no affirmative responses to the Stage One questions and would therefore be considered as having food security; the other 45 percent would typically continue to answer the remaining questions to evaluate the degree of possible food insecurity.[41] Therefore, for the purposes of the economic risk index, we proposed to operationalize the food security items as a binary indicator: yes for those who responded affirmatively to any of the five items, and no for those who did not respond affirmatively to any items. This approach allows all five items to contribute to the scale while using the measure according to the USDA's protocol.

---

[41] These numbers are similar to the latest figures reported by the USDA. According to the report, in 2009, 53.1 percent of households with children who are living at or below 130 percent of the poverty line had food security in 2009 (Nord 2010). (About 82.5 percent of the baseline Baby FACES sample was living at or below 130 percent of the poverty line.)

Based on our decisions for each individual scale, we reviewed three options for creating the final index, each using a low/medium/high scale. Table D.6 presents the proportion of families falling into each category using each classification approach; Table D.7 shows the relationship among the three options by demonstrating how the categorization of individuals would differ in each index.

- Option One classifies individuals with no food risks and up to zero to one financial risk as low-risk; those with a food risk and two financial risks or those with three financial risks and no food risks as medium-risk; and those with three financial risks and a food risk or those with four or five financial risks (regardless of food risks) as high-risk.

- Option Two uses a binary indicator for yes to any of the financial risks and an indicator for any food security risks. Those with "no's" on both scales are at low risk, those with an affirmative on one scale are at medium risk, and those with affirmatives on both scales are at high risk.

- Option Three follows the same principle as option two, but anyone who has been evicted is categorized as high-risk, regardless of his or her other risk factors.

Table D.6. Three Options for Presenting Economic Risk

	Option One	Option Two	Option Three
Low Risk	44.97%	37.21%	37.21%
	No or one financial risk, no food risks	No food or financial risks	No food or financial risks
Medium Risk	43.76%	33.09%	30.67%
	Three financial risks AND no food risks, OR two financial risks, OR at least one food risk and up to two financial risks	At least one food OR at least one financial risk	At least one food OR at least one financial risk, except eviction
High Risk	11.27%	29.70%	32.12%
	Four to five financial risks, OR three financial risks and at least one food risk	At least one food AND at least one financial risk	At least one food AND at least one financial risk, OR eviction

Source: Spring 2009 Parent Interview.

In evaluating the relationships between the different options and other risk measures and child outcomes, we found no pattern of remarkable differences across the three options. For measures of maternal and psychological risk, the pattern was the same regardless of the construction of economic risk–those at higher risk for either psychological risk or maternal risk were, on average, also at a higher level of economic risk. Looking at child outcomes measures, we found no strong relationships between the scales and various child development outcomes, including subscales of the CDI, ASQ-3, and the BITSEA. Between-item correlations were effectively zero for each construction of economic risk, and omnibus F-statistics from ANOVAs were insignificant on each economic risk measure for all but the parent-reported BITSEA Problem domain.[42] For measures with significant F-statistics, additional, pair-wise comparisons revealed patterns that would be expected–those at higher levels of economic risk had higher BITSEA Problem domain scores. For

---

[42] The omnibus F-statistics for the ANOVAs of Option One and the CDI Vocabulary Comprehension raw score was also significant at the 0.05 level. Additional, pair-wise comparisons find a significantly difference only between the means for those at medium risk and those at high risk; those at higher risk have lower CDI Vocabulary Comprehension scores.

Option One, the only significant difference in BITSEA Problem scores among the levels of risk was between low and medium economic risk. For Options Two and Three, however, the means between both low and medium economic risk and low and high economic risk were significantly different, giving slightly greater favor to one of these two options.

Table D.7. Relationship of Three Options for Presenting Economic Risk

Option One	Option Two	Option Three	Percentage	Description of Individuals
Low Risk	Low Risk	Low Risk	37.21	No financial or food risks
Low Risk	Medium Risk	High Risk	1.09	Evicted, but no other financial or food risks
Low Risk	Medium Risk	Medium Risk	6.67	One financial risk except eviction, no food risks
Medium Risk	Medium Risk	Medium Risk	23.40	Two to three financial risks except eviction OR at least one food risk
Medium Risk	Medium Risk	High Risk	0.97	Two to three financial risks and eviction, and no food risks
High Risk	Medium Risk	Medium Risk	0.61	Four financial risks except eviction and no food risks
High Risk	Medium Risk	High Risk	0.36	Four financial risks AND eviction and no food risks
Medium Risk	High Risk	High Risk	19.39	Two to three financial risks, possibly including eviction, AND at least one food risk
High Risk	High Risk	High Risk	10.30	Four or five financial risks, possibly including eviction, AND at least one food risk

Source: Spring 2009 Parent Interview.

We selected Option Two as the final measure. The other options have their merits–Option One weighing financial risks slightly more heavily given the nature of their respective survey instruments, and Option Three assigning more severity of risk to the uniquely dire act of eviction. However, we decided that Option Two best reflects the relationship among the 10 items from the two scales and offers simplicity in its concept. Between-item correlations across the food security and financial security questions are relatively low (particularly between eviction and any of the food security items). The correlations range from effectively zero to 0.33; the highest correlations were between a family's ability to make its housing payments and items related to food running out and conserving resources. Given the lack of any notable relationships among the items on the scales, it seems most appropriate to give the items equal weight, while assigning the most severe risk to those who have a measure of risk on both scales.

Change in the ASQ-3 Administration Mode

We administered the Ages & Stages Questionnaires, Third Edition (ASQ-3) to the 1-year-old Cohort children at age 2 as a measure of children's cognitive, communication, and motor development and to identify children who may be at risk in these developmental areas. We also used this measure in spring 2009 when the children were one year old (see Baby FACES baseline report for the 1-year-old results; ACF 2011). Early Head Start children's ASQ-3 scores fall behind the normative sample at age 1; however, they are catching up with their same-age peers at age 2. One caveat in interpreting the ASQ-3 data for 2-year-olds: A change in the ASQ-3 administration

procedure during data collection from spring 2009 (age 1) to spring 2010 (age 2) might partly contribute to the change in children's scores at age 2.

The design of the ASQ-3 requires advance distribution of the forms to parents prior to face-to-face discussion with interviewers. This step allows parents to see the types of skills or activities assessed from illustrations in the instrument, and affords them the opportunity to try out the skills and activities with their children before completing the forms. In the spring 2009 Baby FACES data collection, we administered the ASQ-3 through telephone interviews. Therefore, parents could not see the items and their illustrations in the forms. We did not send the forms to parents in advance due to the complexities in scheduling the interview and uncertainty of the age form needed by the time of the interview.[43] Parents, therefore, did not know the specific types of skills or abilities included in the ASQ-3 until the time of the interview and thus could have underestimated development at age 1. During spring 2010 Baby FACES data collection, we administered the ASQ-3 through the parent self-administered questionnaire (SAQ) when the interviewers conducted home visits, allowing parents to see the illustrations in the forms and ask the interviewers questions. In addition, we included examples of skills and activities assessed by the ASQ-3 in an advance letter to parents. These parents, therefore, had an opportunity to try out these skills and activities with their children, allowing the assessment to more accurately capture children's general development at age 2.

Early Exiters

To conduct our analysis comparing "early exiters"[44] with continuing Early Head Start participants, we needed to develop a strategy to address levels of missing data from different instruments. For our analysis of the two groups in Chapter VIII, we ran a Student's *t*-test using data from different instruments in our study. These instruments included the program director interview, teacher or home visitor interview, classroom and home visit observations, staff-child report, and parent interview. Since we used different instruments in the analysis of early exiters, we needed to account for both unit and item nonresponse.

We needed to use a method other than applying nonresponse rates to account for both unit and item nonresponse in the instruments. (See Appendix A for description of weights.) Applying our nonresponse weights to address missing reports would have been cumbersome because of the large number of instruments used. Further, it would fail to address any potential biases caused by systematic item nonresponse, because the nonresponse weights account for unit, not item, nonresponse. Instead, we used the same multiple imputation technique that we used in the analysis of the FST data (described above), which allowed us to account for both unit and item nonresponse simultaneously. We created five multiply imputed datasets for the consented population of 973 children and used these data sets in the analysis of early exiters. Box D.2 lists the variables included in our imputation model.

---

[43] Depending on the age of the child on the day of the parent interview, this step required administration of the ASQ-3 10-, 12-, 14-, 16-, or 18-month form.

[44] An "early exiter" is defined as a child who left his or her Early Head Start program before the spring 2010 data collection window and was eligible to receive an exit interview.

Box D.2. Variables in Imputation Model

- **Child demographic and background characteristics**
  - Race/ethnicity
  - Gender
  - Age in months
  - Dual language learner status
  - Health status

- **Family demographic and background characteristics**
  - Mother's age
  - Number of children and adults in the home
  - Single-parent status
  - Household income-to-needs ratio
  - Maternal employment status
  - Maternal education level
  - Teenage mother status
  - Family receives public assistance
  - Family food and financial security issues
  - Mother born in the United States
  - Family moved in the previous year

- **Child outcomes**
  - MacArthur-Bates Communicative Development Inventories (CDI)—Infant Short Form
  - Brief Infant Toddler Social Emotional Assessment (BITSEA)
  - Ages & Stages Questionnaires, Third Edition (ASQ-3)

- **Staff characteristics**
  - Education level
  - Years of experience teaching infants and toddlers
  - Years of experience in Early Head Start
  - Has earned or is working toward degree in early childhood education
  - Depressive symptoms
  - Quality of relationship with child's parents

- **Classroom or home visit quality**
  - Infant Toddler Environment Rating Scale-Revised (ITERS-R)
  - Home Visit Rating Scale-Adapted (HOVRS-R)

- **Family program experiences**
  - Early Head Start service option in spring 2009
  - Quality of relationship with Early Head Start provider
  - Typical number of center days attended or home visits received in the previous year

- **Program characteristics**
  - Total enrollment
  - Staff turnover rate
  - Program morale
  - Number of full- and part-time staff per child
  - Program approach

Exit Interview Analysis

Not only did we need to account for missing reports to compare families who exit Early Head Start programs early with continuing participants, we also needed to address missing information to analyze the parent exit interview. We used exit interview nonresponse weights to account for the exiting families who did not complete an exit interview. Unit-level nonresponse was high: Only 52 percent of exiting families had an exit interview. Appendix A details the methods used to create the nonresponse weights for this analysis.

The item nonresponse on this instrument, however, was quite low. Only one variable that we used in the analysis had missing data, with 3 out of 126 observations missing. Low levels of missing information diminish the potential level of bias due to unavailable data, making multiple imputation unnecessary. In lieu of creating an additional weight to account for nonresponse on this item, we implemented a simpler imputation strategy. The variable with missing values recorded families' satisfaction with their Early Head Start programs on a scale of 1 (very dissatisfied) to 4 (very satisfied). To impute missing values, we created dummy variables for each possible satisfaction level. We then computed the probability that a family in the same cohort and in the same program stratum would have provided a satisfaction-level rating of 1, 2, 3, or 4 in response to the question. We assigned the probability of selecting that satisfaction level to each of the corresponding dummies created for families with missing data in the same cohort and program stratum. For example, for 20 families in the same cohort and program stratum with nonmissing responses, if 5 families gave the satisfaction-level rating of 3, and 15 families answered 4, the probability of answering 3 would be 0.25, and the probability of answering 4 would be 0.75. We would then apply these probabilities to the dummy variables for families with missing data in the same cohort and program stratum: a value of 0.25 to the satisfaction level-3 dummy and the value of 0.75 to the satisfaction level-4 dummy. Assigning a 1, for example, to the level-4 dummy and a 0 to the level-3 dummy would fail to account for uncertainty about the missing responses. The approach described here takes this uncertainty into account.

APPENDIX E. SUPPLEMENTAL TABLES

This appendix provides supplemental tables for Chapter V through Chapter VII. The table numbers indicate which chapter they relate to; for example, tables for Chapter VI are numbered E.VI.1, E.VI.2, and so forth.

While this report focuses on data of the 1-year-old Cohort at age 2, we collected and analyzed data for the Newborn Cohort at age 1 as well. As explained in Appendix D, we compared cohorts at baseline to see if there were any significant differences then as well as when children in both cohorts were 1-year-olds. Since the cohorts are similar at age 1, we are able to combine both cohorts to analyze and describe our full sample at age 1, strengthening the analysis of 1-year-olds participating in the study. Unweighted data for all 1-year-olds in the sample are included in the supplemental tables, presented separately for each cohort as well as for both cohorts combined. We also present all demographic data from Chapter VI for the full 1-year-old Cohort sample at age 2.

Chapter V Supplemental Tables

Newborn and 1-year-old Cohort at Age 1

Table E.V.1 Summary Statistics for Baby FACES Child Care Quality Data at Age 1

Scales	Number of Items	N	Alpha	Unweighted Mean (SD)	Reported Response Range	Possible Response Range
All 1 Year Olds						
ITERS-R Total	32	248	0.88	3.81	1.86-5.84	1 - 7
Personal Care	6	267	0.68	3.10	1.00-6.50	1 - 7
Furnishings	5	276	0.64	3.92	1.60-7.00	1 - 7
Listening and Talking	3	276	0.66	4.29	1.33-7.00	1 - 7
Activities	10	258	0.74	3.51	1.57-6.11	1 - 7
Interaction/Social	4	275	0.72	4.66	1.50-7.00	1 - 7
Program Structure	4	276	0.57	4.19	1.00-7.00	1 - 7
HOVRS-A Overall Quality	7	289	0.84	3.34	1.00-5.00	1 - 5
Visitor Strategies Quality	4	289	0.82	3.18	1.00-5.00	1 - 5
Effectiveness Quality	3	290	0.67	3.56	1.00-5.00	1 - 5
Newborn Cohort at Age 1						
ITERS-R Total	32	43	0.85	3.72	1.97-5.29	1 - 7
Personal Care	6	47	0.73	3.01	1.00-6.50	1 - 7
Furnishings	5	53	0.53	3.80	1.60-5.80	1 - 7
Listening and Talking	3	53	0.28	3.88	2.00-5.67	1 - 7
Activities	10	50	0.55	3.52	1.75-4.90	1 - 7
Interaction/Social	4	52	0.79	4.55	1.75-6.50	1 - 7
Program Structure	4	53	0.59	4.17	1.00-7.00	1 - 7
HOVRS-A Overall Quality	7	49	0.78	3.36	1.86-4.57	1 - 5
Visitor Strategies Quality	4	49	0.80	3.20	1.50-4.75	1 - 5
Effectiveness Quality	3	49	0.58	3.57	2.33-5.00	1 - 5
1-Year-Old Cohort at Age 1						
ITERS-R Total	32	205	0.88	3.83	1.86-5.84	1 - 7
Personal Care	6	220	0.67	3.12	1.17-6.50	1 - 7
Furnishings	5	223	0.65	3.94	1.60-7.00	1 - 7
Listening and Talking	3	223	0.69	4.36	1.33-7.00	1 - 7
Activities	10	208	0.76	3.51	1.57-6.11	1 - 7
Interaction/Social	4	223	0.71	4.68	1.50-7.00	1 - 7
Program Structure	4	223	0.57	4.19	1.33-7.00	1 - 7
HOVRS-A Overall Quality	7	240	0.84	3.34	1.00-5.00	1 - 5
Visitor Strategies Quality	4	240	0.82	3.18	1.00-5.00	1 - 5
Effectiveness Quality	3	241	0.69	3.56	1.00-5.00	1 - 5

Sources: Spring 2009 and 2010 Classroom and Home Visit Observations.

Note: Includes observations of classrooms and home visits to the 1-year-old Cohort in spring 2009 and to the Newborn Cohort in spring 2010.

ITERS-R = Infant/Toddler Environment Rating Scale-Revised.
HOVRS-A = Home Visit Rating Scale-Adapted.

Table E.V.2 Child Care Quality in Early Head Start: Infant/Toddler Environment Rating Scales and Observed Child/Adult Ratios at Age 1

Characteristics	Weighted Mean or Percentage (Standard Error)		
	All 1 Year Olds	Newborn Cohort at Age 1	1-Year-Old Cohort at Age 1
Group Size	5.25 (0.08)	5.53 (0.23)	5.20 (0.09)
Child/Adult Ratio	2.38 (0.04)	2.47 (0.13)	2.36 (0.04)
ITERS-R Total	3.81 (0.04)	3.72 (0.10)	3.83 (0.04)
Personal Care	3.10 (0.06)	3.01 (0.16)	3.12 (0.06)
Furnishings	3.92 (0.05)	3.80 (0.13)	3.94 (0.06)
Listening and Talking	4.29 (0.06)	3.88 (0.11)	4.36 (0.06)
Activities	3.51 (0.05)	3.52 (0.11)	3.51 (0.05)
Interaction/Social	4.66 (0.06)	4.55 (0.16)	4.68 (0.06)
Program Structure	4.19 (0.06)	4.17 (0.18)	4.19 (0.07)
Sample Size	**424-427**		

Sources: Spring 2009 and 2010 Classroom Observations.

Note: Includes observations of classrooms of the 1-year-old Cohort in spring 2009 and of the Newborn Cohort in spring 2010.

ITERS-R = Infant/Toddler Environment Rating Scale-Revised.

Table E.V.3 Characteristics of Early Head Start Home Visits at Age 1

Characteristics	Weighted Mean or Percentage (Standard Error)		
	All 1 Year Olds	Newborn Cohort at Age 1	1-Year-Old Cohort at Age 1
Length of Home Visit (minutes)	83.57 (1.13)	80.21 (2.99)	84.10 (1.22)
Number of Children Participating in Visit	0.72 (0.04)	0.83 (0.12)	0.70 (0.04)
Number of Adults Participating in Visit	1.32 (0.04)	1.09 (0.10)	1.35 (0.05)
Home Visit Conducted in (percentage)			
English	77.28 (2.03)	75.86 (5.67)	77.51 (2.18)
Spanish	28.81 (2.19)	32.76 (6.22)	28.18 (2.35)
Other Language	1.17 (0.52)	0.00 (0.00)	1.36 (0.60)
If Home Visit Conducted in Language Other than English, Interpreter Used (percentage)	9.40 (2.71)	26.32 (10.38)	6.12 (2.43)
Proportion of Home Visit Time per Activity (percentage)			
Child-Focused Activities	48.07 (1.19)	49.74 (3.96)	47.81 (1.23)
Parent/Family-Focused Activities	18.74 (0.82)	18.53 (2.41)	18.77 (0.87)
Parent-Child-Focused Activities	14.24 (0.64)	13.36 (1.71)	14.38 (0.68)
Staff-Family Relationship-Building Activities	13.86 (0.71)	16.14 (1.88)	13.50 (0.77)
Crisis Management Activities	4.89 (0.48)	2.31 (0.82)	5.30 (0.53)
Activities During Home Visit[a] (percentage)			
Child/Parent Observation/Assessment	62.44 (2.35)	51.72 (6.62)	64.13 (2.50)
Evaluation/Feedback on Parent-Child Interactions	41.55 (2.39)	37.93 (6.43)	42.12 (2.58)
Provision of Education and/or Information	69.95 (2.22)	77.59 (5.52)	68.75 (2.42)
Problem Solving	37.56 (2.35)	32.76 (6.22)	38.32 (2.54)
Goal Setting/Planning	50.47 (2.43)	50.00 (6.62)	50.54 (2.61)
Crisis Intervention	8.69 (1.37)	5.17 (2.93)	9.24 (1.51)
Model or Demonstrate Interaction with Child / Facilitate Parent-child Interaction	44.37 (2.41)	41.38 (6.52)	44.84 (2.60)
Observation of Caregiver-Child Interactions	36.62 (2.34)	29.31 (6.03)	37.77 (2.53)
Provision of Emotional Support to Parent	29.81 (2.22)	34.48 (6.30)	29.08 (2.37)
Play	79.11 (1.97)	75.86 (5.67)	79.62 (2.10)
Other	4.69 (1.03)	3.45 (2.42)	4.89 (1.13)
Alignment of Home Visit Activities with Planned Activities[b]	4.29 (0.04)	4.17 (0.12)	4.31 (0.04)
Sample Size	**117-427**		

Source: Spring 2009 and 2010 Home Visit Observations.

Note: Includes observations of home visits to the 1-year-old Cohort in spring 2009 and to the Newborn Cohort in spring 2010.

[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 "very well aligned" with planned activities.

HOVRS-A = Home Visit Rating Scale-Adapted.

Table E.V.4 Observed Home Visit Quality in Early Head Start, Spring 2009 and 2010

Scales	Weighted Mean (Standard Error)		
	All 1 Year Olds	Newborn Cohort at Age 1	1-Year-Old Cohort at Age 1
HOVRS-A Overall Quality	3.34 (0.04)	3.36 (0.08)	3.34 (0.05)
Visitor Strategies Quality	3.18 (0.05)	3.20 (0.10)	3.18 (0.05)
Responsiveness to Family	3.04 (0.06)	3.19 (0.13)	3.02 (0.06)
Relationship with Family	3.92 (0.05)	3.83 (0.12)	3.94 (0.05)
Facilitation of Parent-Child Interaction	2.90(0.06)	2.88 (0.11)	2.90 (0.07)
Non-Intrusiveness	2.85 (0.06)	2.95 (0.13)	2.83 (0.07)
Effectiveness Quality	3.56 (0.04)	3.57 (0.09)	3.56 (0.05)
Parent-Child Interaction	3.23 (0.06)	3.11 (0.15)	3.25 (0.07)
Parent Engagement	3.20 (0.06)	3.24 (0.13)	3.19 (0.06)
Child Engagement	4.27 (0.05)	4.39 (0.10)	4.26 (0.06)
Observer Rating of Visit Quality	3.34 (0.05)	3.26 (0.13)	3.35 (0.05)
Sample Size	**421-427**		

Sources: Spring 2009 and 2010 Home Visit Observations.

Note: Includes observations of home visits to the 1-year-old Cohort in spring 2009 and to the Newborn Cohort in spring 2010.

HOVRS-A = Home Visit Rating Scale-Adapted.

Table E.V.5 Staff-Parent Relationship Quality at Age 1

Characteristic	Weighted Mean (Standard Error)		
	All 1 Year Olds	Newborn Cohort at Age 1	1-Year-Old Cohort at Age 1
Staff-Parent Relationship Quality Score for Children Served in Centers			
Parent Report	4.43 (0.03)	4.55 (0.09)	4.41 (0.03)
Teacher Report	4.13 (0.03)	4.15 (0.10)	4.13 (0.03)
Staff-Parent Relationship Quality Score for Children Receiving Services by Home Visits			
Parent Report	4.59 (0.02)	4.56 (0.09)	4.59 (0.02)
Teacher Report	4.25 (0.03)	4.27 (0.10)	4.24 (0.03)
Sample Size	**860-911**		

Sources: Spring 2009 and 2010 Parent, Teacher, and Home Visitor Interviews.

Note: Includes interviews with parents and staff serving the 1-year-old Cohort in spring 2009 and the the Newborn Cohort in spring 2010.

Chapter VI Supplemental Tables

Newborn and 1-year-old Cohort at Age 1

Table E.VI.1. Household Characteristics of 1 Year Olds (Percentages Unless Otherwise Indicated)

	Total Sample at Age 1	Newborn Cohort at Age 1	1-Year-Old Cohort at Age 1
Child Lives with			
Two biological parents	48.42 (2.50)	43.52 (5.88)	49.19 (2.54)
Married	30.11 (2.13)	29.91 (4.68)	30.15 (2.20)
Unmarried	69.89 (2.13)	70.09 (4.68)	69.85 (2.20)
One biological parent			
Birth mother only	49.30 (2.54)	54.63 (5.84)	48.46 (2.56)
Birth father only	0.63 (0.28)	0.93 (0.93)	0.59 (0.29)
No biological parents	1.64 (0.45)	0.93 (0.94)	1.76 (0.51)
Among Children Living Without Birth Father,			
Child lives with father figure[a]	11.70 (1.76)	7.55 (3.07)	12.46 (1.94)
Mean Number of Adults in Household	1.88 (0.04)	2.06 (0.09)	1.85 (0.04)
Mean Number of Children in Household	2.52 (0.05)	2.60 (0.13)	2.51 (0.06)
Average Household Size	4.42 (0.07)	4.66 (0.18)	4.38 (0.08)
Child Lives in Intergenerational Household	16.73 (1.72)	11.54 (3.11)	17.72 (1.87)
Average Household Income	$24,061 (1823.94)	$17,844 (1287.59)	$25,029 (2104.51)
Median Household Income[b]	$17,500	$15,500	$17,500
Household Income[b]			
$0–$9,999	24.44 (1.59)	34.02 (4.06)	22.95 (1.65)
$10,000–$17,499	25.42 (1.48)	25.77 (4.30)	25.36 (1.61)
$17,500–$24,999	20.28 (1.42)	14.43 (3.31)	21.19 (1.59)
$25,000 or more	29.86 (1.52)	25.77 (4.12)	30.50 (1.67)
Household Income as a Percentage of the Poverty Level[b,c]			
0–50	30.42 (1.73)	47.42 (4.31)	27.77 (1.89)
51–100	39.03 (2.05)	23.71 (4.25)	41.41 (2.30)
101–130[d]	11.25 (1.04)	9.28 (2.60)	11.56 (1.16)
131–higher	19.31 (1.46)	19.59 (3.70)	19.26 (1.59)
Average Number of People Contributing to Household Income	1.59 (0.04)	1.41 (0.06)	1.62 (0.04)
Sample Size	**720-791**	**97-108**	**623-683**

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and 2010 Parent Interview for the Newborn Cohort.

Note: The Newborn Cohort includes pregnant women and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009. For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

[a]Sample size of 53 for Newborn Cohort and 289 for 1-year-old Cohort.

[b]Income-related questions had higher rates of refusal and missing responses than other parent interview variables. There were 11 missing values for income-related questions among Newborn Cohort parents (10 percent missing) and 60 missing values among 1-year-old Cohort parents (9 percent missing).

[c]Poverty level is adjusted for household size according to 2009 and 2010 HHS poverty guidelines.

[d]130 percent of the poverty level is a common eligibility threshold for Head Start, food stamps, and free school lunch. Families over 130 percent of poverty are not eligible to receive Early Head Start services.

Table E.VI.2. Household Characteristics of 1 Year Olds by Child's Race/Ethnicity (Percentages Unless Otherwise Indicated)

	White	African American	Hispanic
Child Lives with			
Two biological parents	51.41 (3.63)	18.28 (3.75)	65.37 (3.86)
Married	36.01 (4.40)	13.99 (2.77)	38.73 (3.74)
Unmarried	63.99 (4.40)	86.01 (2.77)	61.27 (3.74)
One biological parent			
Birth mother only	44.31 (3.72)	79.98 (4.07)	33.51 (3.80)
Birth father only	2.04 (0.98)	0.00 (0.00)	0.00 (0.00)
No biological parents	2.24 (0.99)	1.74 (1.20)	1.11 (0.68)
Among Children Living Without Birth Father,			
Child lives with father figure[a]	19.51 (4.31)	4.30 (2.11)	10.80 (3.76)
Mean Number of Adults in Household	1.93 (0.06)	1.50 (0.08)	2.06 (0.07)
Mean Number of Children in Household	2.36 (0.12)	2.39 (0.13)	2.78 (0.08)
Average Household Size	4.31 (0.15)	3.90 (0.14)	4.90 (0.13)
Child Lives in Intergenerational Household	17.21 (2.54)	16.14 (3.79)	15.35 (2.83)
Average Household Income	$24,474 (1577.11)	$27,777 (6254.70)	$20,785 (1314.09)
Median Household Income[b]	$19,100	$15,000	$12,500
Household Income[b]			
$0–$9,999	20.49 (2.84)	33.36 (2.73)	21.22 (3.00)
$10,000–$17,499	20.23 (2.67)	24.83 (3.63)	28.58 (3.58)
$17,500–$24,999	19.39 (3.48)	16.97 (2.76)	23.55 (3.27)
$25,000 or more	39.88 (3.56)	24.84 (3.15)	26.64 (2.78)
Household Income as a Percentage of the Poverty Level[b,c]			
0–50	23.09 (2.63)	35.54 (3.52)	31.36 (3.04)
51–100	38.30 (3.87)	34.80 (3.57)	41.72 (4.59)
101–130[d]	12.69 (2.29)	10.41 (2.37)	14.28 (2.26)
131–higher	25.92 (3.00)	19.26 (3.47)	12.64 (2.18)
Average Number of People Contributing to Household Income	1.68 (0.07)	1.35 (0.04)	1.55 (0.06)
Sample Size	**228-250**	**136-150**	**257-282**

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and 2010 Parent Interview for the Newborn Cohort.

Note: The Newborn Cohort includes pregnant women and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009. For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

[a]Sample size of 97 White, 101 African-American and 93 Hispanic.

[b]Income-related questions had higher rates of refusal and missing responses than other parent interview variables. There were 11 missing values for income-related questions among Newborn Cohort parents (10 percent missing) and 60 missing values among 1-year-old Cohort parents (9 percent missing).

[c]Poverty level is adjusted for household size according to 2009 and 2010 HHS poverty guidelines.

[d]130 percent of the poverty level is a common eligibility threshold for Head Start, food stamps, and free school lunch. Families over 130 percent of poverty are not eligible to receive Early Head Start services.

Table E.VI.3. Languages Spoken in the Homes of 1 Year Olds (Percentages)

	Total Sample at Age 1	Newborn Cohort at Age 1	1-Year-Old Cohort at Age 1
Language Spoken in the Home			
English only[a]	68.02 (3.23)	69.44 (5.25)	67.79 (3.39)
Spanish (only or primarily)[b]	15.68 (2.15)	11.11 (2.52)	16.40 (2.46)
English (primarily) and Spanish[c]	13.65 (1.76)	18.52 (4.50)	12.88 (1.77)
English (primarily) and other language[d]	1.77 (0.51)	0.93 (0.91)	1.90 (0.54)
Other language (only or primarily)[e]	0.88 (0.44)	0.00 (0.00)	1.02 (0.51)
Among Spanish-Speaking Households, Child Hears Spanish in Household[f]			
All or most of the time	76.92 (3.12)	75.00 (7.37)	77.22 (3.24)
Some of the time or very little	23.08 (3.12)	25.00 (7.37)	22.78 (3.24)
Among Other Language-Speaking Households, Child Hears Other Language in Household[g]			
All or most of the time	44.83 (9.58)	0.00 (0.00)	46.43 (10.02)
Some of the time or very little	55.17 (9.58)	100.00 (0.0)	53.57 (10.02)
Sample Size	**791**	**108**	**683**

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and 2010 Parent Interview for the Newborn Cohort.

Note: The Newborn Cohort includes pregnant women and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009. For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

[a]Includes children to whom parents reported speaking English only at home.

[b]Includes children to whom parents reported speaking Spanish only or most often or who hear Spanish most or all of the time at home.

[c]Poverty level is adjusted for household size according to 2009 and 2010 HHS poverty guidelines.

[d]130 percent of the poverty level is a common eligibility threshold for Head Start, food stamps, and free school lunch. Families over 130 percent of poverty are not eligible to receive Early Head Start services.

[e]Includes children not in the four preceding groups.

[f]Sample size 36 for Newborn Cohort and 237 for 1-year-old Cohort..

[g]Sample size 1 for Newborn Cohort and 28 for 1-year-old Cohort.

Table E.VI.4. Languages Spoken in the Homes of 1 Year Olds by Child's Race/Ethnicity (Percentages)

	White	African American	Hispanic
Language Spoken in the Home			
English only[a]	98.83 (0.69)	93.46 (2.68)	24.82 (3.35)
Spanish (only or primarily)[b]	0.00 (0.00)	0.00 (0.00)	43.13 (4.34)
English (primarily) and Spanish[c]	0.55 (0.54)	3.04 (2.18)	31.39 (3.85)
English (primarily) and other language[d]	0.63 (0.44)	1.99 (1.14)	0.32 (0.32)
Other language (only or primarily)[e]	0.00 (0.00)	1.51 (1.10)	0.34 (0.32)
Among Spanish-Speaking Households, Child Hears Spanish in Household[f]			
All or most of the time	13.37 (12.70)	0.00 (0.00)	83.60 (2.68)
Some of the time or very little	86.63 (12.70)	100.00 (0.00)	16.40 (2.68)
Among Other Language-Speaking Households, Child Hears Other Language in Household[g]			
All or most of the time	15.51 (15.03)	78.25 (9.55)	67.61 (32.64)
Some of the time or very little	84.49 (15.03)	21.75 (9.55)	32.39 (32.64)
Sample Size	**250**	**150**	**228**

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and 2010 Parent Interview for the Newborn Cohort.

Note: The Newborn Cohort includes pregnant women and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009. For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

[a]Includes children to whom parents reported speaking English only at home.

[b]Includes children to whom parents reported speaking Spanish only or most often or who hear Spanish most or all of the time at home.

[c]Poverty level is adjusted for household size according to 2009 and 2010 HHS poverty guidelines.

[d]130 percent of the poverty level is a common eligibility threshold for Head Start, food stamps, and free school lunch. Families over 130 percent of poverty are not eligible to receive Early Head Start services.

[e]Includes children not in the four preceding groups.

[f]Sample size of 11 White, 4 African-American and 229 Hispanic.

[g]Sample size of 5 White, 7 African-American and 13 Hispanic.

Table E.VI.5 Language Spoken to 1 Year Olds by Family Members (Percentages)

	Birth Mother	Birth Father	Grandparent
English	83.83 (2.09)	71.85 (3.40)	84.62 (3.17)
Spanish	30.14 (3.34)	37.83 (4.20)	23.08 (4.47)
Other language	2.85 (0.71)	2.05 (0.95)	1.54 (1.09)
Sample Size	**773**	**341**	**130**

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and 2010 Parent Interview for the Newborn Cohort.

Note: The Newborn Cohort includes pregnant women and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009. For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

[a]Percentages add to more than 100 because each family member could speak to the child in more than one language.

Table E.VI.6 Education and Employment of Mothers of 1 Year Olds (Percentages)

	Total Sample at Age 1	Newborn Cohort at Age 1	1-Year-Old Cohort at Age 1
Highest Education Completed			
Less than high school	40.88 (2.23)	37.96 (5.06)	41.35 (2.26)
High school diploma or equivalent	32.21 (1.74)	37.96 (4.81)	31.28 (1.70)
Some college or AA	22.64 (1.84)	18.52 (4.60)	23.31 (1.93)
BA or higher	4.27 (0.80)	5.56 (2.13)	4.06 (0.78)
Currently Taking Classes	31.06 (2.22)	35.19 (5.59)	30.39 (2.25)
Currently in Job Training	6.85 (0.95)	12.04 (2.99)	6.01 (1.02)
Employed in Last 12 Months	54.39 (2.10)	47.22 (4.60)	55.56 (2.31)
Currently working	39.48 (1.79)	37.96 (4.16)	39.73 (1.94)
Sample Size	**773-776**	**108**	**665-668**

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and 2010 Parent Interview for the Newborn Cohort.

Note: The Newborn Cohort includes pregnant women and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009. For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

Table E.VI.7. Education and Employment of Mothers of 1 Year Olds by Child's Race/Ethnicity (Percentages)

	White	African American	Hispanic
Highest Education Completed			
Less than high school	29.87 (4.08)	38.49 (4.07)	51.17 (3.65)
High school diploma or equivalent	41.19 (4.01)	24.63 (3.70)	27.55 (3.35)
Some college or AA	24.88 (3.58)	33.64 (4.11)	15.41 (2.66)
BA or higher	4.05 (1.98)	3.24 (1.60)	5.87 (1.42)
Currently Taking Classes	25.68 (3.37)	46.58 (4.02)	21.83 (3.58)
Currently in Job Training	4.40 (1.32)	10.94 (2.31)	5.77 (1.64)
Employed in Last 12 Months	60.19 (4.00)	56.86 (3.65)	48.78 (3.94)
Currently working	43.32 (3.21)	41.13 (3.87)	36.70 (3.96)
Sample Size	**242**	**146**	**280**

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and 2010 Parent Interview for the Newborn Cohort.

Note: The Newborn Cohort includes pregnant women and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009. For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

Table E.VI.8. Education and Employment of Fathers of 1 Year Olds (Percentages)

	Total Sample at Age 1	Newborn Cohort at Age 1	1-Year-Old Cohort at Age 1
Highest Education Completed			
Less than high school	44.33 (2.16)	42.39 (5.07)	44.63 (2.34)
High school diploma or equivalent	38.66 (2.08)	43.48 (5.14)	37.92 (2.30)
Some college or AA	12.50 (1.46)	10.87 (3.59)	12.75 (1.48)
BA or higher	4.51 (0.79)	3.26 (1.74)	4.70 (0.91)
Currently Taking Classes	13.45 (1.35)	16.13 (3.75)	13.03 (1.48)
Currently in Job Training	4.67 (0.76)	n.a	4.67 (0.76)
Employed in Last 12 Months	79.94 (1.69)	76.04 (4.76)	80.59 (1.69)
Currently working	62.62 (2.01)	56.25 (4.79)	63.75 (2.11)
Sample Size	**578-688**	**92-96**	**538-596**

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and 2010 Parent Interview for the Newborn Cohort.

Note: The Newborn Cohort includes pregnant women and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009. For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

Table E.VI.9. Education and Employment of Fathers of 1 Year Olds by Child's Race/Ethnicity (Percentages)

	White	African American	Hispanic
Highest Education Completed			
Less than high school	27.78 (3.54)	36.85 (5.07)	59.67 (3.50)
High school diploma or equivalent	50.97 (3.99)	40.91 (5.10)	31.49 (3.63)
Some college or AA	14.48 (2.32)	15.94 (3.64)	7.21 (1.37)
BA or higher	6.78 (1.87)	6.30 (2.15)	1.63 (0.67)
Currently Taking Classes	14.35 (2.75)	12.27 (2.76)	9.53 (2.09)
Currently in Job Training	5.91 (1.87)	3.64 (2.01)	3.38 (1.30)
Employed in Last 12 Months	88.66 (2.64)	62.09 (4.32)	86.09 (2.16)
Currently working	71.58 (3.07)	44.59 (4.76)	71.97(3.39)
Sample Size	**188-223**	**109-128**	**236-246**

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and 2010 Parent Interview for the Newborn Cohort.

Note: The Newborn Cohort includes pregnant women and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009. For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

Table E.VI.10. Financial Difficulties and Public Assistance of Families of 1 Year Olds (Percentages)

Type of Hardship	Total Sample at Age 1	Newborn Cohort at Age 1	1-Year-Old Cohort at Age 1
Could Not Pay the Full Amount of Gas, Oil, or Electricity Bills	31.84 (1.97)	33.64 (4.57)	31.55 (2.08)
Could Not Pay the Full Amount of Rent or Mortgage	30.96 (1.89)	28.97 (4.11)	31.29 (2.03)
Had Service Disconnected by the Telephone Company for Nonpayment	18.92 (1.41)	21.50 (3.68)	18.50 (1.67)
Had Services Turned off by the Gas or Electric Company, or Oil Company Would Not Deliver Oil	9.07 (0.99)	9.35 (2.86)	9.02 (1.07)
Was Evicted from Home or Apartment	5.78 (0.85)	0.93 (0.92)	6.57 (1.01)
Parent Has			
Zero or one financial difficulty	69.25 (1.79)	68.22 (4.17)	69.42 (1.91)
Two or three financial difficulties	25.62 (1.65)	29.91 (3.83)	24.92 (1.76)
Four or five financial difficulties	5.12 (0.80)	1.87 (1.30)	5.66 (0.93)
Family Receives			
WIC	88.80 (1.22)	90.57 (2.95)	88.51 (1.34)
Food stamps	64.38 (2.28)	70.75 (5.07)	63.34 (2.42)
Welfare	30.94 (2.23)	32.08 (4.83)	30.76 (2.35)
SSI	12.93 (1.36)	13.21 (3.24)	12.88 (1.42)
Sample Size	**753-761**	**106-107**	**647-654**

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and 2010 Parent Interview for the Newborn Cohort.

Note: The Newborn Cohort includes pregnant women and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009. For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

SSI = Supplemental Security Income; WIC = Special Supplemental Nutrition Program for Women, Infants, and Children.

Table E.VI.11. Financial Difficulties and Public Assistance of Families of 1 Year Olds by Child's Race/Ethnicity (Percentages)

Type of Hardship	White	African American	Hispanic
Could Not Pay the Full Amount of Gas, Oil, or Electricity Bills	30.96 (2.99)	31.11 (4.36)	35.63 (2.69)
Could Not Pay the Full Amount of Rent or Mortgage	28.40 (2.95)	25.05 (4.17)	34.32 (2.84)
Had Service Disconnected by the Telephone Company for Nonpayment	13.53 (2.51)	18.96 (3.03)	21.48 (2.99)
Had Services Turned off by the Gas or Electric Company, or Oil Company Would Not Deliver Oil	5.65 (1.35)	12.54 (2.60)	9.23 (2.14)
Was Evicted from Home or Apartment	6.94 (1.69)	5.06 (2.07)	4.13 (1.33)
Parent Has			
Zero or one financial difficulty	74.08 (2.75)	76.32 (3.91)	62.75 (2.76)
Two or three financial difficulties	20.56 (2.57)	14.75 (3.09)	34.58 (2.71)
Four or five financial difficulties	5.36 (1.60)	8.93 (2.47)	2.67 (0.99)
Family Receives			
WIC	87.53 (2.51)	85.12 (3.38)	89.63 (2.25)
Food stamps	62.83 (4.26)	74.46 (4.33)	55.41 (4.87)
Welfare	27.78 (4.03)	36.54 (5.38)	31.01 (3.62)
SSI	15.97 (3.09)	13.11 (3.07)	8.82 (2.47)
Sample Size	**237-241**	**145**	**269-271**

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and 2010 Parent Interview for the Newborn Cohort.

Note: The Newborn Cohort includes pregnant women and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009. For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

SSI = Supplemental Security Income; WIC = Special Supplemental Nutrition Program for Women, Infants, and Children.

Table E.VI.12 Food Security of Families of 1 Year Olds (Percentages)

Type of Hardship	Total Sample at Age 1	Newborn Cohort at Age 1	1-Year-Old Cohort at Age 1
Worried Food May Run Out	36.79 (2.03)	29.91 (4.66)	37.92 (2.20)
Food Didn't Last and Didn't Have Money to Get More	28.65 (1.90)	26.17 (3.88)	29.05 (2.05)
Relied on Only a Few Kinds of Low-Cost Foods to Feed Children Because of Financial Reasons	25.46 (1.97)	23.36 (3.60)	25.81 (2.15)
Couldn't Afford to Eat Balanced Meals	24.84 (1.91)	21.50 (3.67)	25.38 (2.04)
Couldn't Feed Children a Balanced Meal for Financial Reasons	17.63 (1.53)	15.89 (3.38)	17.92 (1.72)
Parent Has			
Zero or one food security difficulty	65.79 (2.09)	71.03 (3.88)	64.93 (2.23)
Two or three food security difficulties	16.58 (1.28)	12.15 (3.08)	17.30 (1.46)
Four or five food security difficulties	17.63 (1.62)	16.82 (3.41)	17.76 (1.80)
Sample Size	**758-761**	**107**	**651-654**

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and 2010 Parent Interview for the Newborn Cohort.

Note: The Newborn Cohort includes pregnant women and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009. For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

Table E.VI.13. Food Security of Families of 1 Year Olds by Child's Race/Ethnicity (Percentages)

Type of Hardship	White	African American	Hispanic
Worried Food May Run Out	28.04 (3.32)	29.58 (4.34)	49.28 (3.39)
Food Didn't Last and Didn't Have Money to Get More	19.97 (3.23)	20.52 (3.51)	38.59 (3.36)
Relied on Only a Few Kinds of Low-Cost Foods to Feed			
Children Because of Financial Reasons	15.87 (3.08)	14.68 (3.37)	38.94 (2.82)
Couldn't Afford to Eat Balanced Meals	15.03 (2.83)	19.81 (3.78)	37.44 (3.20)
Couldn't Feed Children a Balanced Meal for Financial			
Reasons	11.67 (2.66)	8.16 (3.23)	31.75 (3.17)
Parent Has			
Zero or one food security difficulty	76.35 (3.53)	76.33 (4.24)	48.99 (3.16)
Two or three food security difficulties	12.89 (2.58)	13.71 (2.89)	22.77 (2.50)
Four or five food security difficulties	10.76 (2.58)	9.95 (2.56)	28.24 (2.80)
Sample Size	**241**	**145**	**271**

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and 2010 Parent Interview for the Newborn Cohort.

Note: The Newborn Cohort includes pregnant women and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009. For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

Table E.VI.14 Living Situation of Families of 1 Year Olds (Percentages Unless Otherwise Indicated)

Living Arrangement	Total Sample at Age 1	Newborn Cohort at Age 1	1-Year-Old Cohort at Age 1
Parent Lives in			
House, apartment, or trailer with family only	91.20 (1.15)	92.52 (2.17)	90.98 (1.23)
House, apartment, or trailer shared with another family	6.57 (0.97)	5.61 (1.77)	6.73 (1.09)
Transitional housing or homeless shelter	1.71 (0.52)	1.87 (1.34)	1.68 (0.57)
Somewhere else	0.53 (0.26)	0.00 (0.00)	0.61 (0.30)
Parent Has Moved in the Past Year	32.19 (1.67)	27.10 (4.37)	33.03 (1.79)
Average Number of Moves[a]	1.43 (0.06)	1.21 (0.09)	1.46 (0.07)
Sample Size	**761**	**107**	**654**

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and 2010 Parent Interview for the Newborn Cohort.

Note: The Newborn Cohort includes pregnant women and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009. For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

[a]Among those who moved.

Table E.VI.15. Living Situation of Families of 1 Year Olds by Child's Race/Ethnicity (Percentages Unless Otherwise Indicated)

Living Arrangement	White	African American	Hispanic
Parent Lives in			
House, apartment, or trailer with family only	93.20 (1.66)	91.42 (3.08)	90.61 (2.13)
House, apartment, or trailer shared with another family	4.00 (1.18)	4.47 (1.63)	8.93 (2.13)
Transitional housing or homeless shelter	1.31 (0.80)	3.18 (1.67)	0.46 (0.34)
Somewhere else	1.48 (0.87)	0.93 (0.89)	0.00 (0.00)
Parent Has Moved in the Past Year	33.43 (3.56)	28.10 (4.12)	28.59 (2.42)
Average Number of Moves[a]	1.40 (0.09)	1.35 (0.10)	1.26 (0.07)
Sample Size	**240**	**145**	**272**

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and 2010 Parent Interview for the Newborn Cohort.

Note: The Newborn Cohort includes pregnant women and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009. For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

[a]Among those who moved.

Table E.VI.16 Economic Risk of Parents of 1 Year Olds (Percentages)[a]

Parent Has	Total Sample at Age 1	Newborn Cohort at Age 1	1-Year-Old Cohort at Age 1
Lower economic risk	37.58 (2.16)	36.45 (5.26)	37.77 (2.29)
Medium economic risk	33.38 (1.80)	34.58 (4.59)	33.18 (1.89)
Highest economic risk	29.04 (1.88)	28.97 (4.22)	29.05 (1.97)
Sample Size	**761**	**107**	**654**

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and 2010 Parent Interview for the Newborn Cohort

Note: The Newborn Cohort includes pregnant women and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009. For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

[a]Economic risk is an index that aggregates financial difficulties and food security difficulties. Parents with fewer than two financial difficulties and fewer than two food security difficulties were classified as at lower economic risk. Parents with two or three financial difficulties or two or three food security difficulties were classified as at medium economic risk. Parents with at least four difficulties in either category were classified as at highest economic risk.

Table E.VI.17. Economic Risk of Parents of 1 Year Olds by Child's Race/Ethnicity (Percentages)[a]

Parent Has	White	African American	Hispanic
Lower economic risk	46.99 (3.83)	40.80 (3.53)	29.30 (2.72)
Medium economic risk	30.89 (4.09)	35.62 (3.18)	31.16 (3.11)
Highest economic risk	22.13 (2.97)	23.58 (4.02)	39.54 (2.76)
Sample Size	**241**	**145**	**271**

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and 2010 Parent Interview for the Newborn Cohort

Note: The Newborn Cohort includes pregnant women and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009. For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

[a]Economic risk is an index that aggregates financial difficulties and food security difficulties. Parents with fewer than two financial difficulties and fewer than two food security difficulties were classified as at lower economic risk. Parents with two or three financial difficulties or two or three food security difficulties were classified as at medium economic risk. Parents with at least four difficulties in either category were classified as at highest economic risk.

Table E.VI.18 Demographic Risk of Mothers of 1 Year Olds (Percentages)

Risk Factor	Total Sample at Age 1	Newborn Cohort at Age 1	1-Year-Old Cohort at Age 1
Single Mother	47.09 (2.35)	53.70 (6.01)	46.04 (2.32)
Teenage Mother	53.06 (2.43)	54.55 (5.34)	52.82 (2.51)
No High School Credential	40.54 (2.14)	37.96 (5.06)	40.96 (2.18)
Receive Public Assistance	72.09 (2.20)	78.30 (4.15)	71.08 (2.35)
Not Employed, in School, or in Training	37.61 (2.38)	38.89 (5.28)	37.41 (2.46)
Maternal Demographic Risk Index[a]			
0–2 (lower risk)	50.19 (2.00)	46.73 (5.36)	50.75 (2.03)
3 (medium risk)	29.09 (1.62)	31.78 (5.22)	28.66 (1.70)
4–5 (highest risk)	20.72 (1.71)	21.50 (4.18)	20.60 (1.70)
Sample Size	**756-790**	**106-110**	**650-682**

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and 2010 Parent Interview for the Newborn Cohort.

Note: The Newborn Cohort includes pregnant women and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009. For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

[a]This index was constructed by summing the five risk factors enumerated in the top portion of the table.

Table E.VI.19. Demographic Risk of Mothers of 1 Year Olds by Child's Race/Ethnicity (Percentages)

Risk Factor	White	African American	Hispanic
Single Mother	39.45 (3.03)	78.59 (4.00)	31.74 (3.65)
Teenage Mother	48.10 (3.39)	58.68 (5.16)	54.37 (3.39)
No High School Credential	28.24 (3.79)	37.69 (3.95)	51.39 (3.68)
Receive Public Assistance	71.59 (4.42)	80.50 (4.07)	63.64 (4.12)
Not Employed, in School, or in Training	36.04 (3.57)	27.49 (3.35)	46.40 (4.22)
Maternal Demographic Risk Index[a]			
0–2 (lower risk)	60.49 (4.30)	39.86 (4.20)	49.48 (3.32)
3 (medium risk)	26.74 (3.54)	31.51 (4.22)	29.08 (2.98)
4–5 (highest risk)	12.77 (2.21)	28.63 (4.24)	21.44 (3.18)
Sample Size	**236-250**	**145-150**	**271-282**

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and 2010 Parent Interview for the Newborn Cohort.

Note: The Newborn Cohort includes pregnant women and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009. For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

[a]This index was constructed by summing the five risk factors enumerated in the top portion of the table.

Table E.VI.20. Health Care Services and Health Status of Families of 1 Year Olds in the Past Year

	Total Sample at Age 1	Newborn Cohort at Age 1	1-Year-Old Cohort at Age 1
Parent's health status (mean rating)	2.38 (0.04)	2.40 (0.11)	2.38 (0.04)
Percentage of parents in excellent or very good health	52.12 (1.82)	53.27 (4.83)	51.93 (1.96)
Percentage of parents in fair or poor health	12.96 (1.22)	18.69 (3.77)	12.02 (1.28)
Percentage of families that have a regular health care provider	80.03 (1.44)	80.37 (3.84)	79.97 (1.55)
Percentage of families in which at least one family member visited a dentist	79.94 (1.55)	n.a	79.94 (1.55)
Family's health insurance status[a]			
A private health insurance plan	44.81 (1.81)	50.96 (4.90)	43.83 (1.95)
A public/government insurance	78.84 (1.48)	83.18 (3.62)	78.13 (1.62)
No health insurance	7.36 (0.95)	3.74 (1.83)	7.95 (1.06)
Early Head Start has helped find insurance for those who have it (percentage)	23.18 (3.45)	36.36 (10.26)	20.93 (3.59)
Family member needs health care but couldn't obtain it (percentage)	8.04 (0.99)	11.32 (3.08)	7.50 (1.03)
Sample Size	**668-776**	**104-107**	**648-669**

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and 2010 Parent Interview for the Newborn Cohort.

Note: The Newborn Cohort includes pregnant women and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009. For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

[a]The estimates are not mutually exclusive and therefore sum to more than the estimated percentage of parents with coverage.

Table E.VI.21. Parenting Beliefs for Parents of 1 Year Olds

	Total Sample at Age 1	Newborn Cohort at Age 1	1-Year-Old Cohort at Age 1
Parental Modernity Scale			
Traditional Attitudes		19.89 (0.34)	19.78 (0.14)
Progressive Attitudes	20.11 (0.12)	20.36 (0.33)	20.07 (0.14)
Parent spanked the child in the past week	11.44 (1.15)	8.41 (2.68)	11.93 (1.26)
Sample Size	**755-769**	**107**	**648-662**

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and 2010 Parent Interview for the Newborn Cohort.

Note: The Newborn Cohort includes pregnant women and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009. For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

Table E.VI.22. Family Routines for Parents of 1 Year Olds

	Total Sample at Age 1	Newborn Cohort at Age 1	1-Year-Old Cohort at Age 1
Number of days per week family eats dinner together	5.39 (0.07)	5.46 (0.18)	5.38(0.08)
Percentage of families that eat dinner together			
Every day	49.15 (1.81)	49.06 (4.86)	49.17 (1.95)
Five or six days a week	22.35 (1.51)	23.58 (4.13)	22.15 (1.62)
One or two days a week	7.71 (0.97)	7.55 (2.57)	7.74 (1.04)
Child is fed at regular times in a typical day	94.52 (0.82)	94.39 (2.23)	94.55 (0.88)
Number of meals child eats in a typical day	3.30 (0.02)	3.48 (0.08)	3.28 (0.02)
Number of snacks child eats in a typical day	2.47 (0.03)	2.35 (0.09)	2.49 (0.04)
Number of times child wakes up during night	0.83 (0.04)	0.91 (0.08)	0.81 (0.04)
Percentage of children who			
Sleep through the night	45.86 (1.81)	36.45 (4.66)	47.40 (1.95)
Wake up once	34.69 (1.73)	42.06 (4.78)	33.49 (1.85)
Wake up twice or more	19.45 (1.44)	21.50 (3.97)	19.11 (1.54)
Number of hours child sleeps per night	9.42 (0.07)	9.11 (0.18)	9.48 (0.08)
Number of naps child takes in a typical day	1.68 (0.03)	1.83 (0.07)	1.65 (0.03)
Number of hours child naps	1.53 (0.03)	n.a	1.53 (0.03)
Sample Size	**647-767**	**101-107**	**623-660**

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and 2010 Parent Interview for the Newborn Cohort.

Note: The Newborn Cohort includes pregnant women and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009. For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

Table E.VI.23. Child Safety Practices of Families of 1 Year Olds

	Total Sample at Age 1	Newborn Cohort at Age 1	1-Year-Old Cohort at Age 1
Family uses car seat for child	99.67 (0.24)	100.00 (0.00)	99.59 (0.29)
Family uses a gate or door at the top of stairs	96.49 (0.75)	99.07 (0.92)	95.93 (0.89)
Family's home has working smoke alarms	95.34 (0.86)	95.37 (2.02)	95.33 (0.95)
Family has covers on electrical outlets that child can reach	91.18 (1.16)	87.96 (3.13)	91.89 (1.23)
Family uses guards or gates for windows	29.05 (1.86)	28.70 (4.36)	29.12 (2.05)
Sample Size	**599-601**	**108**	**491-493**

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and 2010 Parent Interview for the Newborn Cohort.

Note: The Newborn Cohort includes pregnant women and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009. For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

Table E.VI.24 Mental Health of Parents of 1 Year Olds

	Total Sample at Age 1	Newborn Cohort at Age 1	1-Year-Old Cohort at Age 1
Parent's Mental Health			
CESD-SF Raw Score	5.15 (0.20)	4.78 (0.56)	5.21 (0.22)
CESD-SF: moderate to severe depressive symptoms (percentage)	15.92 (1.33)	14.95 (3.45)	16.08 (1.44)
CESD-SF: no or mild depressive symptoms (percentage)	84.08 (1.33)	85.05 (3.45)	83.92 (1.44)
PSI: Parental Distress raw score	10.78 (0.17)	10.28 (0.49)	10.86 (0.18)
PSI: Parent-Child Dysfunctional Interaction raw score	8.65 (0.15)	7.82 (0.31)	8.79 (0.16)
Parent Substance Use			
Smoking inside the home[b]	17.79 (2.41)	17.65 (6.54)	17.81 (2.59)
Currently smoking	21.29 (1.48)	23.36 (4.09)	20.95 (1.59)
Currently drinking	21.90 (1.62)	n.a	21.90 (1.62)
Drug use in the past year	2.30 (0.59)	n.a	2.30 (0.59)
Ever had a drinking or drug problem[a]	5.53 (0.83)	0.00 (0.00)	6.43 (0.96)
Parents Received (percentages)			
Any health treatment	19.84 (1.45)	20.56 (3.91)	19.72 (1.56)
Treatment for an emotional, personal, or mental problem	16.95 (1.36)	19.63 (3.84)	16.51 (1.45)
Treatment for a drug or alcohol problem	6.04 (0.86)	2.80 (1.60)	6.57 (0.97)
Percentage Early Head Start Helped to Get the Treatment for Those Who Received Treatment[c]	23.18 (3.45)	36.36 (10.26)	20.93 (3.59)
Sample Size	**653-761**	**107**	**649-654**

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and 2010 Parent Interview for the Newborn Cohort.

Note: CESD-SF severe depressive symptoms are defined as scores of 15 or higher; moderate depressive symptoms as scores of 10 or higher but lower than 15; mild depressive symptoms as scores of 5 or higher but lower than 10; no depressive symptoms as scores lower than 5.

The Newborn Cohort includes pregnant women and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009. For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

[a]Parents in Newborn Cohort were asked in 2010 if they had had a drinking or drug problem in the past year.

[b]N=253, 34, 219

[c]N = 151,22,129

PSI = Parenting Stress Index; CESD-SF = Center for Epidemiologic Studies Depression Scale Short Form

Table E.VI.25 Family Functioning for Parents of 1 Year Olds

	Total Sample at Age 1	Newborn Cohort at Age 1	1-Year-Old Cohort at Age 1
FES-Family Conflict raw score	1.29 (0.05)	1.29 (0.05)	n.a
Percentage with FES score over 2	8.00 (2.73)	8.00 (2.71)	n.a
Sample Size	**100**	**100**	

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and 2010 Parent Interview for the Newborn Cohort.

Note: The Newborn Cohort includes pregnant women and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009. For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

FES = Family Environment Scale.

1-year-old Cohort at Age 2

Table E.VI.26. Household Characteristics of 2 Year Olds (Percentages Unless Otherwise Indicated)

	1-Year-Old Cohort at Age 2
Child Lives with	
Two biological parents	54.73 (2.99)
Married	35.80 (2.77)
Unmarried	64.20 (2.77)
One biological parent	
Birth mother only	41.29 (2.91)
Birth father only	0.55 (0.42)
No biological parents	3.43 (0.92)
Among Children Living Without Birth Father,	
Child lives with father figure[a]	18.72 (3.34)
Mean Number of Adults in Household	2.01 (0.05)
Mean Number of Children in Household	2.69 (0.09)
Average Household Size	4.70 (0.11)
Child Lives in Intergenerational Household	6.76 (1.30)
Average Household Income	$22,267 (738.72)
Median Household Income[b]	$18,100
Household Income[b]	
$0–$9,999	19.94 (2.12)
$10,000–$17,499	23.25 (2.46)
$17,500–$24,999	22.76 (2.84)
$25,000 or more	34.04 (2.19)
Household Income as a Percentage of the Poverty Level[b,c]	
0–50	28.49 (2.78)
51–100	35.29 (2.83)
101–130[d]	18.27 (2.06)
131–higher	17.96 (1.97)
Average Number of People Contributing to Household Income	1.47 (0.04)
Sample Size	**434-475**

Source: Spring 2010 Parent Interview.

Note: The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010.

[a]Sample size of 203 for 1-year-old Cohort.

[b]Income-related questions had higher rates of refusal and missing responses than other parent interview variables. There and 41 missing values among 1-year-old Cohort parents (9 percent missing).

[c]Poverty level is adjusted for household size according to 2010 HHS poverty guidelines.

[d]130 percent of the poverty level is a common eligibility threshold for Head Start, food stamps, and free school lunch. Families over 130 percent of poverty are not eligible to receive Early Head Start services.

Table E.VI.27. Household Characteristics of 2 Year Olds by Child's Race/Ethnicity (Percentages Unless Otherwise Indicated)

	White	African American	Hispanic
Child Lives with			
Two biological parents	61.34 (4.82)	28.09 (4.87)	66.05 (4.24)
Married	37.02 (4.82)	17.57 (4.09)	45.16 (3.69)
Unmarried	62.98 (4.82)	82.43 (4.09)	54.84 (3.69)
One biological parent			
Birth mother only	33.50 (4.80)	68.95 (4.71)	31.04 (4.06)
Birth father only	1.12 (1.11)	0.00 (0.00)	0.40 (0.40)
No biological parents	4.03 (1.71)	2.96 (1.42)	2.51 (1.63)
Among Children Living Without Birth Father,			
Child lives with father figure[a]	32.17 (6.76)	4.99 (2.91)	18.52 (6.25)
Mean Number of Adults in Household	2.06 (0.07)	1.74 (0.12)	2.13 (0.11)
Mean Number of Children in Household	2.51 (0.20)	2.50 (0.13)	2.95 (0.09)
Average Household Size	4.57 (0.20)	4.24 (0.17)	5.08 (0.17)
Child Lives in Intergenerational Household	5.87 (1.71)	6.30 (2.38)	4.33 (1.91)
Average Household Income	$25,273 (1857.64)	$18,652 (1650.59)	$21,893 (1064.93)
Median Household Income[b]	$22,000	$14,400	$18,000
Household Income[b]			
$0–$9,999	16.32 (3.47)	29.99 (4.66)	15.22 (3.20)
$10,000–$17,499	15.93 (2.73)	29.01 (5.70)	28.66 (4.13)
$17,500–$24,999	29.85 (4.95)	15.68 (3.85)	23.42 (4.59)
$25,000 or more	37.90 (5.33)	25.32 (3.92)	32.70 (3.23)
Household Income as a Percentage of the Poverty Level[b,c]			
0–50	20.26 (4.04)	37.79 (5.64)	29.64 (4.80)
51–100	33.80 (4.28)	32.66 (5.17)	40.27 (4.54)
101–130[d]	25.33 (3.84)	15.09 (4.41)	13.96 (3.08)
131–higher	20.60 (3.73)	14.46 (3.57)	16.13 (3.58)
Average Number of People Contributing to Household Income	1.51 (0.07)	1.36 (0.07)	1.52 (0.08)
Sample Size	**134-150**	**89-93**	**179-188**

Source: Spring 2010 Parent Interview.

Note: The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010.

[a]Sample size of 57 for white, 62 for African American and 60 for Hispanic, .

[b]Income-related questions had higher rates of refusal and missing responses than other parent interview variables. There were 16 missing values for income-related questions among white parents (11 percent missing), 6 missing values among African American parents (6 percent missing), and 15 missing values among Hispanic parents (8 percent missing).

[c]Poverty level is adjusted for household size according to 2009 and 2010 HHS poverty guidelines.

[d]130 percent of the poverty level is a common eligibility threshold for Head Start, food stamps, and free school lunch. Families over 130 percent of poverty are not eligible to receive Early Head Start services.

Table E.VI.28. Languages Spoken in the Homes of 2 Year Olds (Percentages)

	1-Year-Old Cohort at Age 2
Language Spoken in the Home	
English only[a]	65.51 (3.95)
Spanish (only or primarily)[b]	16.22 (3.24)
English (primarily) and Spanish	13.92 (2.21)
English (primarily) and other language	3.69 (1.43)
Other language (only or primarily)[c]	0.66 (0.66)
Among Spanish-Speaking Households, Child Hears Spanish in Household[d]	
All or most of the time	76.54 (3.90)
Some of the time or very little	23.46 (3.90)
Among Other Language-Speaking Households, Child Hears Other Language in Household[e]	
All or most of the time	44.72 (14.67)
Some of the time or very little	55.28 (14.67)
Sample Size	**475**

Source: Spring 2010 Parent Interview.

Note: The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010.

[a]Includes children to whom parents reported speaking English only at home.

[b]Includes children to whom parents reported speaking Spanish only or most often or who hear Spanish most or all of the time at home.

[c]Includes children not in the four preceding groups.

[d]Sample size equals 160.

[e]Sample size equals 24.

Table EVI.29. Languages Spoken in the Homes of 2 Year Olds by Child's Race/Ethnicity (Percentages)

	White	African American	Hispanic
Language Spoken in the Home			
English only[a]	93.52 (2.40)	96.59 (1.67)	23.61 (4.66)
Spanish (only or primarily)[b]	0.47 (0.47)	0.00 (0.00)	43.18 (5.93)
English (primarily) and Spanish[c]	3.22 (1.38)	0.00 (0.00)	33.21 (4.43)
English (primarily) and other language[d]	2.33 (1.91)	2.42 (1.39)	0.00 (0.00)
Other language (only or primarily)[e]	0.47 (0.46)	0.99 (1.00)	0.00 (0.00)
Among Spanish-Speaking Households, Child Hears Spanish in Household[f]			
All or most of the time	79.77 (12.80)	0.00 (0.00)	19.24 (3.67)
Some of the time or very little	20.23 (12.80)	0.00 (0.00)	80.76 (3.67)
Among Other Language-Speaking Households, Child Hears Other Language in Household[g]			
All or most of the time	12.55 (11.64)	37.70 (21.69)	0.00 (0.00)
Some of the time or very little	87.45 (11.64)	62.30 (21.69)	100.00 (0.00)
Sample Size	**150**	**93**	**188**

Source: Spring 2010 Parent Interview.

Note: The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010.

[a]Includes children to whom parents reported speaking English only at home.

[b]Includes children to whom parents reported speaking Spanish only or most often or who hear Spanish most or all of the time at home.

[e]Includes children not in the four preceding groups.

[f]Sample size equals 8 for White children, 0 for African American children and 149 for Hispanic Children

[g]Sample size equals 5 for White children, 6 for African American children and 2 for Hispanic Children.

Table E.VI.30. Language Spoken to 2 Year Olds by Family Members (Percentages)

	Birth Mother	Birth Father	Grandparent
English	82.81 (2.86)	100.00 (0.00)	100.00 (0.00)
Spanish	31.82 (4.21)	25.36 (6.59)	0.00 (0.00)
Other language	5.00 (1.62)	26.55 (1.02)	16.47 (14.91)
Sample Size	**443**	**11**	**11**

Source: Spring 2010 Parent Interview.

Note: The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010.

[a]Percentages add to more than 100 because each family member could speak to the child in more than one language.

Table E.VI.31. Education and Employment of Mothers of 2 Year Olds (Percentages)

	1-Year-Old Cohort at Age 2
Highest Education Completed	
Less than high school	36.01 (2.88)
High school diploma or equivalent	34.59 (2.96)
Some college or AA	23.22 (2.40)
BA or higher	6.17 (1.11)
Currently Taking Classes	31.45 (2.82)
Currently in Job Training	7.57 (1.35)
Employed in Last 12 Months	58.49 (3.03)
Currently working	44.00 (3.04)
Sample Size	**456-462**

Source: Spring 2010 Parent Interview.

Note: The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010.

Table E.VI.32. Education and Employment of Mothers of 2 Year Olds by Child's Race/Ethnicity (Percentages)

	White	African American	Hispanic
Highest Education Completed			
Less than high school	26.99 (4.75)	37.95 (4.99)	48.31 (4.06)
High school diploma or equivalent	42.52 (5.54)	25.32 (4.77)	28.96 (3.96)
Some college or AA	25.15 (3.93)	30.41 (4.17)	16.97 (3.64)
BA or higher	5.34 (2.21)	6.31 (2.73)	5.76 (1.74)
Currently Taking Classes	32.97 (4.38)	42.85 (6.21)	23.45 (2.78)
Currently in Job Training	7.73 (2.56)	10.39 (4.26)	6.15 (2.03)
Employed in Last 12 Months	58.25 (5.40)	68.21 (4.08)	52.23 (4.81)
Currently working	41.55 (5.53)	50.64 (5.04)	43.51 (3.83)
Sample Size	**143-146**	**90**	**182**

Source: Spring 2010 Parent Interview.

Note: The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010.

Table E.VI.33. Education and Employment of Fathers of 2 Year Olds (Percentages)

	1-Year-Old Cohort at Age 2
Highest Education Completed	
Less than high school	42.72 (3.06)
High school diploma or equivalent	38.36 (2.83)
Some college or AA	12.40 (1.63)
BA or higher	6.52 (1.50)
Currently Taking Classes	12.24 (1.96)
Currently in Job Training	
Employed in Last 12 Months	77.70 (2.75)
Currently working	66.17 (2.96)
Sample Size	**381-417**

Source: Spring 2010 Parent Interview.

Note: The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010.

Table E.VI.34. Education and Employment of Fathers of 2 Year Olds by Child's Race/Ethnicity (Percentages)

	White	African American	Hispanic
Highest Education Completed			
Less than high school	31.47 (4.64)	35.23 (5.57)	58.12 (4.35)
High school diploma or equivalent	48.08 (4.67)	36.96 (6.58)	29.79 (4.36)
Some college or AA	13.30 (2.81)	15.14 (4.60)	9.00 (2.21)
BA or higher	7.15 (2.79)	12.66 (5.24)	3.09 (1.21)
Currently Taking Classes	8.73 (2.70)	16.08 (5.07)	12.62 (2.85)
Currently in Job Training			
Employed in Last 12 Months	75.53 (4.26)	60.97 (6.14)	89.42 (2.68)
Currently working	62.45 (4.82)	51.54 (5.66)	78.88 (3.53)
Sample Size	**123-135**	**69-78**	**154-165**

Source: Spring 2010 Parent Interview.

Note: The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010.

Table E.VI.35. Financial Difficulties and Public Assistance of Families of 2 Year Olds (Percentages)

Type of Hardship	1-Year-Old Cohort at Age 2
Could Not Pay the Full Amount of Gas, Oil, or Electricity Bills	36.52 (2.64)
Could Not Pay the Full Amount of Rent or Mortgage	32.45 (2.93)
Had Service Disconnected by the Telephone Company for Nonpayment	19.39 (2.35)
Had Services Turned off by the Gas or Electric Company, or Oil Company Would Not Deliver Oil	12.25 (2.25)
Was Evicted from Home or Apartment	5.09 (1.07)
Parent Has	
Zero or one financial difficulty	68.35 (2.95)
Two or three financial difficulties	24.95 (2.52)
Four or five financial difficulties	6.70 (1.27)
Family Receives	
WIC	79.54 (2.42)
Food stamps	66.77 (3.09)
Welfare	26.31 (2.87)
SSI	15.52 (2.36)
Sample Size	**453-459**

Source: Spring 2010 Parent Interview.

Note: The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010.

SSI = Supplemental Security Income; WIC = Special Supplemental Nutrition Program for Women, Infants, and Children.

Table E.VI.36. Financial Difficulties and Public Assistance of Families of 2 Year Olds by Child's Race/Ethnicity (Percentages)

Type of Hardship	White	African American	Hispanic
Could Not Pay the Full Amount of Gas, Oil, or Electricity Bills	37.62 (4.38)	38.66 (7.20)	38.11 (3.48)
Could Not Pay the Full Amount of Rent or Mortgage	33.17 (5.46)	29.85 (6.85)	33.51 (4.46)
Had Service Disconnected by the Telephone Company for Nonpayment	16.42 (3.05)	21.49 (3.83)	19.53 (4.18)
Had Services Turned off by the Gas or Electric Company, or Oil Company Would Not Deliver Oil	13.91 (3.81)	19.02 (5.92)	10.57 (2.44)
Was Evicted from Home or Apartment	5.04 (1.98)	2.03 (1.46)	6.22 (2.23)
Parent Has			
Zero or one financial difficulty	68.43 (4.83)	64.94 (7.25)	70.40 (3.73)
Two or three financial difficulties	24.54 (4.30)	25.47 (5.12)	24.11 (3.52)
Four or five financial difficulties	7.03 (2.21)	9.60 (3.55)	5.49 (1.88)
Family Receives			
WIC	73.24 (4.47)	72.04 (5.41)	90.13 (2.57)
Food stamps	69.86 (4.82)	78.91 (5.77)	59.45 (5.13)
Welfare	23.59 (4.25)	36.69 (5.44)	25.34 (3.27)
SSI	21.49 (4.20)	13.04 (3.96)	10.39 (2.69)
Sample Size	**144**	**90**	**182**

Source: Spring 2010 Parent Interview.

Note: The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010.

SSI = Supplemental Security Income; WIC = Special Supplemental Nutrition Program for Women, Infants, and Children.

Table E.VI.37. Food Security of Families of 2 Year Olds (Percentages)

Type of Hardship	1-Year-Old Cohort at Age 2
Worried Food May Run Out	32.65 (2.68)
Food Didn't Last and Didn't Have Money to Get More	26.35 (2.62)
Relied on Only a Few Kinds of Low-Cost Foods to Feed Children Because for Financial Reasons	23.91 (2.31)
Couldn't Afford to Eat Balanced Meals	21.68 (2.19)
Couldn't Feed Children a Balanced Meal for Financial Reasons	17.90 (1.84)
Parent Has	
Zero or one food security difficulty	69.27 (2.41)
Two or three food security difficulties	13.19 (1.83)
Four or five food security difficulties	17.54 (2.04)
Sample Size	**457**

Source: Spring 2010 Parent Interview.

Note: The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010.

Table E.VI.38. Food Security of Families of 2 Year Olds by Child's Race/Ethnicity (Percentages)

Type of Hardship	White	African American	Hispanic
Worried Food May Run Out	22.57 (4.98)	18.68 (4.90)	42.82 (3.51)
Food Didn't Last and Didn't Have Money to Get More	19.28 (4.02)	17.89 (4.86)	36.16 (3.76)
Relied on Only a Few Kinds of Low-Cost Foods to Feed Children Because of Financial Reasons	16.72 (3.67)	15.20 (4.49)	32.48 (3.52)
Couldn't Afford to Eat Balanced Meals	14.09 (3.74)	15.54 (4.62)	33.55 (3.44)
Couldn't Feed Children a Balanced Meal for Financial Reasons	10.71 (2.88)	17.04 (4.89)	26.57 (2.46)
Parent Has			
Zero or one food security difficulty	78.40 (4.25)	79.17 (5.00)	59.61 (3.27)
Two or three food security difficulties	10.18 (3.28)	10.35 (3.23)	12.88 (2.48)
Four or five food security difficulties	11.42 (3.24)	10.48 (4.31)	27.51 (3.56)
Sample Size	**143**	**90**	**181**

Source: Spring 2010 Parent Interview.

Note: The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010.

Table E.VI.39. Living Situation of Families of 2 Year Olds (Percentages Unless Otherwise Indicated)

Living Arrangement	1-Year-Old Cohort at Age 2
Parent Lives in	
House, apartment, or trailer with family only	92.68 (1.25)
House, apartment, or trailer shared with another family	5.29 (1.22)
Transitional housing or homeless shelter	0.59 (0.33)
Somewhere else	1.44 (0.45)
Parent Has Moved in the Past Year	29.21 (2.46)
Average Number of Moves[a]	1.39 (0.07)
Sample Size	**457**

Source: Spring 2010 Parent Interview.

Note: The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010.

[a]Among those who moved.

Table E.VI.40. Living Situation of Families of 2 Year Olds by Child's Race/Ethnicity (Percentages Unless Otherwise Indicated)

Living Arrangement	White	African American	Hispanic
Parent Lives in			
House, apartment, or trailer with family only	93.18 (2.32)	91.05 (3.49)	91.65 (2.81)
House, apartment, or trailer shared with another family	5.17 (2.07)	6.93 (3.09)	6.23 (2.12)
Transitional housing or homeless shelter	0.49 (0.48)	0.00 (0.00)	0.24 (0.25)
Somewhere else	1.16 (1.14)	2.02 (1.99)	1.88 (1.38)
Parent Has Moved in the Past Year	28.98 (4.21)	24.15 (4.85)	30.49 (3.50)
Average Number of Moves[a]	1.52 (0.14)	1.53 (0.21)	1.18 (0.06)
Sample Size	**143**	**90**	**181**

Source: Spring 2010 Parent Interview.

Note: The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010.

[a]Among those who moved.

Table E.VI.41. Economic Risk of Families of 2 Year Olds (Percentages)[a]

Parent Has	1-Year-Old Cohort at Age 2
Lower economic risk	39.95 (2.47)
Medium economic risk	31.26 (2.57)
Highest economic risk	28.80 (2.74)
Sample Size	**459**

Source: Spring 2010 Parent Interview.

Note: The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010.

[a]Economic risk is an index that aggregates financial difficulties and food security difficulties. Parents with fewer than two financial difficulties and fewer than two food security difficulties were classified as at lower economic risk. Parents with two or three financial difficulties or two or three food security difficulties were classified as at medium economic risk. Parents with at least four difficulties in either category were classified as at highest economic risk.

Table E.VI.42. Economic Risk of Parents of 2 Year Olds by Child's Race/Ethnicity (Percentages)[a]

Parent Has	White	African American	Hispanic
Lower economic risk	46.38 (4.37)	43.33 (6.29)	33.01 (3.02)
Medium economic risk	32.22 (4.75)	36.96 (5.94)	28.49 (3.90)
Highest economic risk	21.40 (4.57)	19.72 (5.24)	38.51 (4.01)
Sample Size	**144**	**90**	**182**

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and 2010 Parent Interview for the Newborn Cohort

Note: The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010.

[a]Economic risk is an index that aggregates financial difficulties and food security difficulties. Parents with fewer than two financial difficulties and fewer than two food security difficulties were classified as at lower economic risk. Parents with two or three financial difficulties or two or three food security difficulties were classified as at medium economic risk. Parents with at least four difficulties in either category were classified as at highest economic risk.

Table E.VI.43. Demographic Risk of Mothers of 2 Year Olds (Percentages)

Risk Factor	1-Year-Old Cohort at Age 2
Single Mother	38.54 (2.70)
Teenage Mother	48.73 (2.77)
No High School Credential	34.79 (2.89)
Receive Public Assistance	72.70 (3.15)
Not Employed, in School, or in Training	37.33 (2.97)
Maternal Demographic Risk Index[a]	
0–2 (lower risk)	57.71 (3.34)
3 (medium risk)	25.64 (2.80)
4–5 (highest risk)	16.64 (2.09)
Sample Size	**457-558**

Source: Spring 2010 Parent Interview.

Note: The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010.

[a]This index was constructed by summing the five risk factors enumerated in the top portion of the table.

Table E.VI.44. Demographic Risk of Mothers of 2 Year Olds by Child's Race/Ethnicity (Percentages)

Risk Factor	White	African American	Hispanic
Single Mother	27.72 (4.25)	67.64 (5.10)	29.99 (3.77)
Teenage Mother	46.12 (4.57)	59.65 (5.81)	50.32 (3.98)
No High School Credential	25.53 (4.64)	37.95 (4.99)	46.95 (4.20)
Receive Public Assistance	77.39 (4.57)	85.22 (4.16)	64.27 (4.95)
Not Employed, in School, or in Training	36.98 (5.13)	29.66 (5.82)	41.95 (3.78)
Maternal Demographic Risk Index[a]			
0–2 (lower risk)	66.64 (4.52)	41.43 (5.76)	55.45 (5.41)
3 (medium risk)	19.18 (3.20)	26.87 (5.03)	30.13 (5.04)
4–5 (highest risk)	14.18 (2.99)	31.70 (4.55)	14.42 (3.59)
Sample Size	**145-185**	**89-106**	**183-215**

Source: Spring 2010 Parent Interview.

Note: The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010.

[a]This index was constructed by summing the five risk factors enumerated in the top portion of the table.

Table E.VI.45. Family Routines for Parents of 2 Year Olds

	1-Year-Old Cohort at Age 2
Number of days per week family eats dinner together	5.77 (0.08)
Percentage of families that eat dinner together	
Every day	58.20 (2.44)
Five or six days a week	19.77 (2.18)
One or two days a week	6.23 (1.15)
Child is fed at regular times in a typical day	96.34 (0.93)
Number of meals child eats in a typical day	3.22 (0.03)
Number of snacks child eats in a typical day	2.54 (0.06)
Number of times child wakes up during night	0.59 (0.05)
Percentage of children who	
Sleep through the night	58.12 (2.72)
Wake up once	29.47 (2.70)
Wake up twice or more	12.41 (2.12)
Number of hours child sleeps per night	9.65 (0.11)
Number of naps child takes in a typical day	1.08 (0.02)
Number of hours child naps	
Sample Size	**459-461**

Source: Spring 2010 Parent Interview.

Note: The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010.

Table E.VI.46. Child Safety Practices of Families of 2 Year Olds

	1-Year-Old Cohort at Age 2
Family uses car seat for child	99.09 (0.49)
Family uses a gate or door at the top of stairs	98.93 (0.57)
Family's home has working smoke alarms	96.88 (0.86)
Family has covers on electrical outlets that child can reach	86.61 (1.72)
Family uses guards or gates for windows	25.04 (2.24)
Sample Size	**466-467**

Source: Spring 2010 Parent Interview.

Note: The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010.

Table E.VI.47. Family Functioning for Parents of 2 Year Olds

	1-Year-Old Cohort at Age 2
FES-Family Conflict raw score	1.28 (0.02)
Percentage with FES score over 2	4.53 (1.22)
Sample Size	**429**

Source: Spring 2010 Parent Interview.

Note: The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010.

FES = Family Environment Scale.

Chapter VII Supplemental Tables

Newborn and 1-year-old Cohort at Age 1

Table E.VII.1. Child Health status and Health Care Access at Age 1

Child Characteristics	Unweighted Means or Percentages (SE)		
	Total Sample at Age 1	Newborn Cohort at Age 1	1-Year-Old Cohort at Age 1
Child Born Premature	8.66 (1.03)	4.76 (2.08)	9.29 (1.14)
Birth Weight			
Percentage low birth weight	6.57 (0.90)	4.67 (2.04)	6.88 (0.99)
Percentage very low birth weight	1.31 (0.41)	0.00 (0.00)	1.53 (0.48)
Health Status	1.78 (0.03)	1.76 (0.09)	1.78 (0.03)
Percentage Excellent or Very Good Health	78.61 (1.49)	75.93 (4.12)	79.05 (1.59)
Percentage Fair or Poor Health	5.38 (0.82)	5.56 (2.21)	5.35 (0.88)
Percentage Have Regular Health Care Provider	97.67 (0.54)	98.13 (1.31)	97.59 (0.59)
Percentage Received Any Health Services	99.74 (0.19)	100.00 (0.00)	99.69 (0.22)
Child Has Ever Visited			
A doctor for a check-up (percentage)	100.00 (0.00)	100.00 (0.00)	100.00 (0.00)
A dentist (percentage)	24.05 (1.55)	28.04 (4.34)	23.39 (1.66)
Child's Last Regular Doctor Check-Up Was Fewer than 6 Months Ago (percentage)	97.89 (0.52)	94.39 (2.23)	98.47 (0.48)
Frequency of Well-Baby Check-Ups (percentage)			
Never	1.05 (0.37)	2.83 (1.61)	0.76 (0.34)
Once or twice	5.64 (0.84)	8.49 (2.71)	5.18 (0.86)
3–4 times	22.54 (1.51)	35.85 (4.66)	20.40 (1.57)
5–9 times	44.82 (1.80)	44.34 (4.83)	44.90 (1.94)
10 times or more	25.95 (1.59)	8.49 (2.71)	28.77 (1.77)
Percentage Have Sufficient Well-Child Doctor Visits	70.67 (1.66)	52.83 (4.85)	73.60 (1.74)
Child's Immunization Status Is "Completely Up to Date" (percentage)	91.13 (1.04)	87.85 (3.16)	91.67 (1.09)
Percentage Ever Been Hospitalized	18.65 (1.39)	16.67 (3.59)	18.97 (1.50)
Child's Health Insurance Status[a,] (percentage)			
A private health insurance plan	34.44 (1.73)	39.42 (4.79)	33.64 (1.86)
A public/government insurance	87.39 (1.20)	92.52 (2.54)	86.54 (1.34)
No health insurance	3.68 (0.68)	2.80 (1.60)	3.82 (0.75)
Child has dental insurance coverage	51.28 (1.83)	74.29 (4.27)	47.50 (1.98)
Early Head Start helped to find health insurance[b]	11.08 (1.16)	15.38 (3.54)	10.37 (1.22)
Sample Size			
Parent Interview	**745-788**	**105-108**	**640-680**

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and spring 2010 Parent Interview for the Newborn Cohort.

Note: The Newborn Cohort includes pregnant women (N = 85) and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 (including 69 who enrolled during pregnancy with that child). For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

We defined prematurity as having been born more than three weeks preterm, low birth weight as weighing less than 2.5 kg, and very low birth weight as weighing less than 1.5 kg.

[a] The estimates are not mutually exclusive and hence sum to more than the estimated percentage of children with coverage.

[b] Only for those with no health insurance (N = 731,104, 627 for the total sample, Newborn Cohort, and 1-year-old Cohort, respectively).

Table E.VII.2 Child General Development (Parent-Reported) at Age 1: ASQ-3

Outcome	Total Sample at Age 1	Newborn Cohort at Age 1	1-Year-Old Cohort at Age 1
	Unweighted Means or Percentages (Standard Error)		
ASQ-3 Raw Score			
Communication	40.90 (0.49)	44.53 (1.16)	40.32 (0.54)
Gross Motor	50.00 (0.58)	47.10 (1.36)	50.67 (0.64)
Fine Motor	43.72 (0.55)	45.09 (1.23)	43.40 (0.61)
Problem Solving	40.61 (0.59)	42.23 (1.31)	40.23 (0.66)
Personal-Social	42.82 (0.53)	42.24 (1.12)	42.96 (0.60)
Total Score	217.17 (2.09)	221.21 (4.41)	216.23 (2.36)
ASQ Cut-Off Score (2SDs below the mean or lower)			
Communication	6.53 (0.88)	2.80 (1.60)	7.12 (0.99)
Gross Motor	11.29 (1.33)	14.02 (3.36)	10.65(1.44)
Fine Motor	14.64 (1.49)	18.69 (3.77)	13.70 (1.60)
Problem Solving	19.05 (1.65)	14.95 (3.45)	20.00 (1.87)
Personal-Social	8.47 (1.17)	7.48 (2.54)	8.70 (1.31)
ASQ in the monitoring zone (1-2SDs below the mean)			
Communication	21.77 (1.48)	16.82 (3.62)	22.55 (1.61)
Gross Motor	9.70 (1.24)	15.89 (3.54)	8.26 (1.28)
Fine Motor	17.46 (1.60)	17.76 (3.70)	17.39 (1.77)
Problem Solving	20.46 (1.70)	16.82 (3.62)	21.30 (1.91)
Personal-Social	22.75 (1.76)	18.69 (3.77)	23.70 (1.98)
Sample Size			
Parent Interview	**567-781**	**107**	**460-674**

Source: Spring 2009 Parent Interview for the 1-year-old Cohort and 2010 Parent Interview for the Newborn Cohort.

Note: The Newborn Cohort includes pregnant women (N = 85) and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 (including 69 who enrolled during pregnancy with that child). For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

ASQ-3 = Ages & Stages Questionnaires (Third Edition); SD = standard deviation.

Table E.VII.3. Children's Vocabulary Development (Staff-Reported) at Age 1: CDI

CDI Scores	Unweighted Means or Percentages (Standard Error)		
	Total Sample at Age 1	Newborn Cohort at Age 1	1-Year-Old Cohort at Age 1
CDI English Raw Score			
Vocabulary Comprehension	30.12 (0.73)	29.29 (1.95)	30.26 (0.79)
Vocabulary Production	2.65 (0.19)	2.13 (0.39)	2.74 (0.22)
CDI Spanish Raw Score (the EHSREP Form)			
Vocabulary Comprehension	35.20 (1.93)	31.45 (4.61)	35.86 (2.11)
Vocabulary Production	2.14 (0.31)	2.05 (0.64)	2.16 (0.34)
CDI Conceptual Score (English and Spanish)			
Vocabulary Comprehension	41.08 (1.84)	39.16 (4.93)	41.37 (1.98)
Vocabulary Production	3.05 (0.40)	3.21 (1.06)	3.02 (0.43)
CDI Spanish Raw Score (the Officially Published Form)[a]			
Vocabulary Comprehension	30.21 (4.50)	30.21 (4.39)	n.a
Vocabulary Production	2.95 (1.02)	2.95 (0.99)	n.a
Sample Size			
SCR English CDI	**810**	**119**	**691**
SCR Spanish CDI	**133-142**	**19-20**	**113-123**

Source: Spring 2009 Staff Child Report (SCR) for the 1-year-old Cohort and 2010 SCR for the Newborn Cohort.

Note: The Newborn Cohort includes pregnant women (N = 85) and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 (including 69 who enrolled during pregnancy with that child). For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010.

[a] Only used in 2010 for the newborn cohort.

CDI = MacArthur-Bates Communicative Development Inventories.

Table E.VII.4. Child Social Emotional Development at Age 1

Outcome	Unweighted Means or Percentages (Standard Error)		
	Total Sample at Age 1	Newborn Cohort at Age 1	1-Year-Old Cohort at Age 1
Parent-Reported BITSEA Raw Score			
Problem Domain	10.97 (0.33)	13.48 (1.81)	10.57 (0.24)
Competence Domain	16.38 (0.20)	17.74 (1.22)	16.16 (0.13)
Staff-Reported BITSEA Raw Score			
Problem Domain	6.19 (0.17)	6.26 (0.51)	6.18 (0.18)
Competence Domain	12.80 (0.13)	12.88 (0.32)	12.79 (0.14)
Parent-Reported BITSEA Cutoff Score			
Problem Domain	0.27 (0.02)	0.30 (0.04)	0.27 (0.02)
Competence Domain	0.10 (0.01)	0.11 (0.03)	0.10 (0.01)
Staff-Reported BITSEA Cutoff Score			
Problem Domain	0.14 (0.01)	0.14 (0.03)	0.14 (0.01)
Competence Domain	0.14 (0.01)	0.13 (0.03)	0.15 (0.01)
Parent-Reported BITSEA Screening Positive (percentage)	0.33 (0.02)	0.35 (0.05)	0.33 (0.02)
Staff-Reported BITSEA Screening Positive (percentage)	0.24 (0.02)	0.23 (0.04)	0.25 (0.02)
Sample Size			
** Parent Interview**	**781-787**	**108**	**673-679**
** Staff Child Report**	**746-768**	**118-122**	**628-646**

Source: Spring 2009 Parent Interview and Staff Child Report (SCR) for the 1-year-old Cohort and 2010 Parent Interview and SCR for the Newborn Cohort.

Note: The Newborn Cohort includes pregnant women (N = 85) and babies up to 8 weeks of age who were enrolled in Early Head Start in spring 2009 and continued to be enrolled in spring 2010 when the children were 1. The 1-year-old Cohort includes families of children between 10 and 15 months of age who were enrolled in Early Head Start in spring 2009 (including 69 who enrolled during pregnancy with that child). For tables that describe 1-year-olds, we include data from the 1-year-old Cohort collected in spring 2009 and from the still enrolled Newborn Cohort in spring 2010. BITSEA = Brief Infant-Toddler Social & Emotional Assessment.

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