Developmental Psychology
2005, Vol. 41, No. 6, 885–901

Copyright 2005 by the American Psychological Association
0012-1649/05/$12.00 DOI: 10.1037/0012-1649.41.6.885

The Effectiveness of Early Head Start for 3-Year-Old Children and Their Parents: Lessons for Policy and Programs

John M. Love, Ellen Eliason Kisker, and Christine Ross
Mathematica Policy Research, Inc.

Jill Constantine and Kimberly Boller
Mathematica Policy Research, Inc.

Rachel Chazan-Cohen
U.S. Department of Health and Human Services

Christy Brady-Smith and Allison Sidle Fuligni
Columbia University

Helen Raikes
Society for Research in Child Development Visiting Scholar, U.S. Department of Health and Human Services

Jeanne Brooks-Gunn
Columbia University

Louisa Banks Tarullo
U.S. Department of Health and Human Services

Peter Z. Schochet, Diane Paulsell, and Cheri Vogel
Mathematica Policy Research, Inc.

Early Head Start, a federal program begun in 1995 for low-income pregnant women and families with infants and toddlers, was evaluated through a randomized trial of 3,001 families in 17 programs. Interviews with primary caregivers, child assessments, and observations of parent–child interactions were completed when children were 3 years old. Caregivers were diverse in race–ethnicity, language, and other characteristics. Regression-adjusted impact analyses showed that 3-year-old program children performed better than did control children in cognitive and language development, displayed higher emotional engagement of the parent and sustained attention with play objects, and were lower in aggressive behavior. Compared with controls, Early Head Start parents were more emotionally supportive, provided more language and learning stimulation, read to their children more, and spanked less. The strongest and most numerous impacts were for programs that offered a mix of home-visiting and center-based services and that fully implemented the performance standards early.

Keywords: early childhood development, poverty in early childhood, child policy and intervention, Head Start

John M. Love, Ellen Eliason Kisker, Christine Ross, Jill Constantine, Kimberly Boller, Peter Z. Schochet, Diane Paulsell, and Cheri Vogel, Mathematica Policy Research, Inc., Princeton, NJ; Helen Raikes, Society for Research in Child Development Visiting Scholar, Administration for Children and Families; Rachel Chazan-Cohen, and Louisa Banks Tarullo, Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services; Jeanne Brooks-Gunn, Teachers College and College of Physicians and Surgeons, Columbia University; Christy Brady-Smith and Allison Sidle Fuligni, Teachers College, Columbia University.

Ellen Eliason Kisker is now at Twin Peaks Consulting, LLC, Longmont, Colorado; Allison Sidle Fuligni is now at the University of California, Los Angeles, Center for Improving Child Care Quality, Graduate School of Education and Information Studies; Helen Raikes is now at the Department of Family and Consumer Sciences, University of Nebraska—Lincoln. Louisa Banks Tarullo is now at Mathematica Policy Research, Inc., Princeton, New Jersey.

The findings reported here are based on research conducted as part of the National Early Head Start Research and Evaluation Project funded by the Administration for Children and Families (ACF), U. S. Department of Health and Human Services under Contract 105-95-1936 to Mathematica Policy Research, Princeton, NJ, and Columbia University's National Center for Children and Families, Teachers College, in conjunction with the Early Head Start Research Consortium. The consortium consists of representatives from 17 programs participating in the evaluation, 15 local research teams, the evaluation contractors, and ACF. Research institutions in the consortium (and principal researchers conducting this research with children through 36 months of age) include ACF (Rachel Chazan-Cohen, Judith Jerald, Esther Kresh, and Helen Raikes); Catholic University of America (Michaela Farber, Harriet Liebow, Nancy Taylor, Elizabeth Timberlake, and Shavaun Wall); Columbia University (Lisa Berlin, Christy Brady-Smith, Jeanne Brooks-Gunn, and Allison Sidle Fuligni); Harvard University (Catherine Ayoub, Barbara Alexander Pan, and Catherine Snow); Iowa State University (Dee Draper, Gayle Luze, Susan McBride, and Carla Peterson); Mathematica Policy Research (Kimberly Boller, Jill Constantine, Ellen Eliason Kisker, John M. Love, Diane Paulsell, Christine Ross, Peter Schochet, Susan Sprachman, Louisa Banks Tarullo, Cheri Vogel, and Welmoet van Kammen); Medical University of South Carolina (Richard Faldowski, Gui-Young Hong, and Susan Pickrel); Michigan State University (Hiram Fitzgerald, Tom Reischl, and Rachel Schiffman); New York University (Mark Spellmann and Catherine Tamis-LeMonda); University of Arkansas (Robert Bradley, Richard Clubb, Andrea Hart, Mark Swanson, and Leanne Whiteside-Mansell); University of California, Los Angeles (Carollee Howes and Claire Hamilton); University of Colorado Health Sciences Center (Robert Emde, Jon Korfmacher, JoAnn Robinson, Paul Spicer, and Norman Watt); University of Kansas (Jane Atwater, Judith Carta; and Jean Ann Summers); University of Missouri—Columbia (Mark Fine, Jean Ispa, and Kathy Thornburg); University of Pittsburgh (Beth Green, Carol McAllister, and Robert McCall); University of Washington School of Education (Eduardo Armijo and Joseph Stowitschek); University of Washington School of Nursing (Kathryn Barnard and Susan Spieker), and Utah State University (Lisa Boyce, Gina Cook, Catherine Callow-Heusser, and Lori Roggman).

The content of this publication does not necessarily reflect the views or policies of the U.S. Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

We owe a special debt to the late Helen Taylor, deputy commissioner for Head Start during the formative years of Early Head Start, for her vision of strong research in the service of understanding programs. We thank the staff of the 17 participating Early Head Start programs that recruited families and supported the random assignment process and the more than 3,000 families who participated in the extensive interviews, observations, and assessments. We thank the data collection coordinating staff of Mathematica Policy Research led by Ms. Sprachman and Dr. van Kammen and the 17 teams of interviewers and assessors supervised by the local research teams, who participated as part of the Early Head Start Research Consortium. We acknowledge the videotape coders at Columbia University under the supervision of Dr. Brooks-Gunn. Anne Bloomenthal of Mathematica Policy Research prepared all data files for analysis.

Correspondence concerning this article should be addressed to John M. Love, Mathematica Policy Research, Inc. P.O. Box 2393, Princeton, NJ 08543. E-mail: jlove@mathematica-mpr.com

Since its inception in 1995, Early Head Start has become a major national initiative. It began with 68 grantees funded by the Administration on Children, Youth, and Families (ACYF) in the U.S. Department of Health and Human Services (USDHHS), and it has grown to more than 700 programs that, in 2004, served 62,000 low-income families with infants and toddlers throughout the country. Early Head Start is a comprehensive, two-generation program that focuses on enhancing children's development while strengthening families. The programs, designed to serve low-income pregnant women and families with infants and toddlers up to age 3, use many strategies to provide a wide range of services.[1] All programs are required to follow the Head Start Program Performance Standards, which stipulate that programs are to provide high-quality, comprehensive child development services delivered through home visits, child care, case management, parenting education, health care and referrals, and family support (USDHHS, 1995). Programs formally select a program model—home-based, center-based, or a combination of the two (a "mixed" approach)—after completing a community resources and needs assessment. As is true of all Head Start programs, federal monitors visit once every 3 years to assess adherence to the performance standards.

Study Rationale and Research Questions

This article summarizes the impacts of Early Head Start on child and parent outcomes near the end of participation in the program when the children were 3 years old, by addressing three primary research questions.[2] Our expectations are informed by three decades of theory, research, and program evaluations. Early childhood intervention programs are based on the premise that it is possible to alter outcomes such as cognitive, emotional, and social skills in young children (Bloom, 1964; Bronfenbrenner, 1979; Brooks-Gunn, 2004). Developmental perspectives consider continuity and change over time and the circumstances and events that promote them over the life course (Rutter, 2000). This perspective is valuable both theoretically and practically. Theoretically, we want to know whether developmental trajectories can be altered, by how much, for which children, under what circumstances, and at what ages. Practically, we want to know what programs or strategies can enhance developmental trajectories. Intervention programs seek to alter development in a positive direction for children at risk for poor outcomes (Brooks-Gunn, 2003; Guralnick, 1997). Early childhood interventions typically provide direct services to the child and intervene with families, with the hope that changes in parents will result in changes for children or for both parents and children (Brooks-Gunn & Markman, 2005). Some of the most compelling evidence has come from two-generation programs, showing that trajectories of at-risk children can be altered by providing stimulating, stable, and caring environments for them (Bornstein, 2002; Meisels & Shonkoff, 2000; Shonkoff & Phillips, 2000).

At the same time, these demonstrations of efficacy have left many questions unanswered. These questions have to do with (a) the generalizability of findings from early childhood programs studied with only small, homogeneous samples; (b) the ability of federally funded and nationally implemented community-based programs to alter child development trajectories (in contrast to single-site demonstration programs implemented by individual investigators); (c) the role of fidelity in implementation of the program model; and (d) the issue of whether the intervention should target parents, children, or both.

This study addresses each of these issues by assessing a federally funded, two-generation initiative that was implemented nationwide. The programs enrolled low-income families that were highly diverse in race–ethnicity; age (both teenage and older mothers); having first- and later-born children; and living in urban, suburban, or small town areas. We systematically assessed the extent to which programs implemented the federal performance standards, and studied programs that adopted home-based, center-based, and mixed approaches to delivering services. The study's research questions and the rationale for each follow:

1. Did the Early Head Start programs have significant impacts on child and parenting outcomes at age 3, when the program ended?

Based on previous evaluations of interventions with very young children, we expected to find significant treatment effects on cognitive and language outcomes (Barnett, 1995; Farran, 2000; Karoly et al., 1998; Yoshikawa, 1995). However, we expected somewhat smaller impacts than those found in previous studies because the age at which children entered and left their Early Head Start program varied (programs had latitude as to when to enroll families—prenatally or during the first year of life),[3] and the average length of enrollment for Early Head Start families in the research sample was 22 months. Thus, families received a somewhat shorter intervention than those that participated in studies that have found impacts with large effect sizes on cognitive and language development at age 3. Two such studies were the Infant Health and Development Program (IHDP), in which enrollment began at birth and continued to age 3 (Brooks-Gunn, Klebanov, Liaw, & Spiker, 1993; IHDP, 1990) and the Carolina Abecedarian project, a demonstration program in which services were consistently intensive beginning when children were around 4 months old and continuing until elementary school enrollment (Campbell & Ramey, 1994; Ramey & Campbell, 1984).

The evaluation of the Comprehensive Child Development Program (CCDP), a forerunner of Early Head Start also funded by ACYF, found significant impacts on children's cognitive development at age 2 that largely disappeared by age 3 (St. Pierre, Layzer, Goodson, & Bernstein, 1997). We expected that Early Head Start would yield larger impacts than the CCDP, however, because of its focus on child development in contrast to the CCDP's more general case management approach (St. Pierre & Layzer, 1999).

Previous evaluations of infant–toddler programs have produced mixed findings regarding impacts on reducing behavior problems, so we expected to find fewer or weaker impacts on this outcome. For example, the literature shows increased parent-reported behavior problems in the Abecedarian study (Campbell & Ramey, 1994)

[1] Early Head Start programs are part of the federal Head Start program. An Early Head Start grantee does not have to be an agency that also administers a Head Start program, however.

[2] The evaluation's technical report (ACF, 2002a) contains findings related to additional research questions.

[3] For the evaluation, programs could enroll families with children up to age 12 months. The average age at random assignment was 5 months, with one quarter of the families enrolling while the mother was still pregnant with the focus child.

but lowered rates of problem behaviors in other studies (IHDP, 1990; Johnson & Walker, 1987; and Lally, Mangione, & Honig, 1988), but see Yoshikawa (1995) for a review of long-term impacts. Few studies have examined observed social–emotional outcomes, such as engagement and attention, although the IHDP reported positive impacts on persistence, enthusiasm, and overall involvement during a play task when children were 30 months old (Spiker, Ferguson, & Brooks-Gunn, 1993).

Parenting outcomes were important to Early Head Start programs as well. Through site visits that researchers conducted with the 17 Early Head Start research programs early in their implementation, we learned that programs often stressed parent–child relationship goals because of the expected indirect effects they would have on children's development (Administration for Children and Families [ACF], 2002b). Similarly, theories of change in the early childhood intervention literature often view the parent–child relationship and family home environment as key agents of change in the child's life (Shonkoff & Phillips, 2000). Many intervention programs that begin in the child's first year or earlier report positive effects on mother–child interaction or parenting. In a comprehensive review of early childhood program effects on parents, 17 of the 20 studies that examined parenting or parent–child interaction among home-based, center-based, or mixed-approach programs reported favorable impacts in such areas as responsiveness, sensitivity, attachment, warmth, and discipline practices (Brooks-Gunn & Markman, 2005; Brooks-Gunn, Berlin, & Fuligni, 2000). The literature is less clear regarding program impacts on negative aspects of parenting. Therefore, expectations about the impacts of Early Head Start on detachment were less certain than expectations that Early Head Start would lead to more responsive parenting.

2. Does adherence to the Head Start Program Performance Standards matter?

Despite the often stated need to assess whether programs being evaluated are implemented as designed (Gomby, 1999; Wang & Reynolds, 2000), no large-scale study has measured program implementation and accounted for it in analyses of program impacts. Although program evaluations are sometimes criticized for examining programs too early in their development, before they become fully implemented (Gilliam, Ripple, Zigler, & Leiter, 2000; McCall & Green, 2004; McCall, Ryan, & Plemons, 2003), the Early Head Start evaluation studied 17 programs from the first two waves of 143 programs to be funded. Unlike other evaluations, however, it was designed to include extensive site visits and well-developed criteria for assessing implementation. The implementation study allowed us to classify programs according to their pattern of implementation during the evaluation period (ACF, 2002b) and to analyze impacts by those patterns—about one third of the programs fully implemented the performance standards early in their development, one third did so later, and the rest did not do so during the evaluation period. We expected to find a greater number of or larger impacts on all outcome measures in those programs that were fully implemented as early as possible during the evaluation period (Garrett, McKinney, Kinukawa, Redd, & Moore, 2003).

3. Are impacts more likely to be found, or likely to be greater in magnitude, in Early Head Start programs that offer both center- and home-based services (a mixed approach) than in programs that offer only, or primarily, home- or center-based services?

The evaluation of Early Head Start is unique in testing whether diverse programs serving varied populations according to federal performance standards can meet the needs of their communities by selecting locally appropriate program approaches (center-based, home-based, or mixed). We anticipated that adopting a mixed approach, combining home- and center-based services, might be more effective than adopting only one of these approaches. Programs combining home visitation with center-based care have been effective across a wider range of outcomes than programs using either approach alone (Brooks-Gunn et al., 1993; Wasik, Ramey, Bryant, & Sparling, 1990). Center-based programs and programs offering combined services (such as Project CARE) are more likely to produce significant impacts on children's cognitive development (Barnett, 1995; Benasich, Brooks-Gunn, & Clewell, 1992; Campbell & Ramey, 1994; Karoly et al.,1998; Ramey & Ramey, 1998; Wasik et al., 1990). Although the majority of home visiting programs have not found cognitive impacts, some do, although particularly in single-site programs (Sweet & Appelbaum, 2004).[4] The majority of studies of home-based and combined programs have found modest effects on parenting and parent outcomes (Brooks-Gunn et al., 2000; Daro & Harding, 1999; Dokecki, Hargrave, & Sandler, 1983; Gomby, 1999; Kitzman et al., 2000; McCarton et al., 1997; Nauta & Travers, 1982; Olds et al., 1999; St. Pierre et al., 1997).

Method

Participants and Procedures

The 17 research programs were located in all regions of the country, in both urban and rural settings, and included all major Early Head Start program approaches.[5] Early Head Start programs recruit families with incomes at or below the federal poverty level and are required to use at least 10% of available spaces to serve children with disabilities. Up to 10% of program spaces may also be used to enroll families with incomes above the poverty level. To be eligible for enrollment in the research programs, families had to meet the program's income guidelines, agree to random assignment, and be expecting a child or have a child under 12 months of

[4] Early Head Start home-visiting sites were not included in this meta-analysis of exclusively home-visiting programs.

[5] As a condition of funding, the first 68 Early Head Start programs agreed to participate in the evaluation, along with a local research partner, if selected. In March 1996, 41 university research teams submitted proposals—together with nearby programs—to conduct local research and participate in the national evaluation. ACYF purposively selected 15 research sites to provide a national geographic distribution reflecting the major programmatic approaches and settings and the diverse family characteristics typical of Early Head Start families nationally. The strength of the local research team was also a factor in the selection of research sites. Each program had to be able to recruit twice as many families (typically, 150) as it could serve. The initial selection resulted in fewer center-based programs than desired, so ACYF selected 2 additional ones, for a full sample of 17 (all of which were funded in 1995 and 1996, the first 2 years of Early Head Start program funding). Programs also agreed to enroll children before 12 months of age only. (Early Head Start children can be enrolled up to age 3, although programs are encouraged to enroll children in early infancy or during the prenatal period.)

age. Seventeen Early Head Start programs across the country recruited 3,001 families to participate in the evaluation. Except for recruiting twice as many families as they could serve, programs were expected to recruit the same way they would have done in the absence of the research, with special instructions to include all types of families they were designed to serve (including those whose babies had disabilities). Although programs could apply to have families excused from participating in the research, no exemptions were requested.

After programs determined that families met eligibility guidelines, they sent the names to Mathematica Policy Research, which randomly assigned families to the program (n = 1,513) or control group (n = 1,488). In a few sites where programs had concerns about balancing assignments across key types of families (such as families of different racial–ethnic backgrounds or families living in different service areas), eligible applicants were stratified according to their key characteristics before random assignment. Control group families could not receive Early Head Start services, but they could access other services in the community.

Sample enrollment, baseline data collection, and random assignment began in July 1996 and were completed in September 1998. Program staff collected baseline data from all families during the application and enrollment process, before randomization. Random assignment yielded equivalent groups, as seen in the highly similar baseline characteristics of program and control group members (see Table 1).

Sample Characteristics

Primary caregivers of the focus children (mothers in 99% of the families) were diverse (see Table 1 for both family and child characteristics at baseline), including whether the child enrolled in the program had been born yet (24% of the program mothers were pregnant); mother's age (39% were under 20 years old, 33% were 20 to 25, and 28% were older than 25); parity (62% of children were firstborn); mother's education (48% had less than a 12th-grade education); race–ethnicity (37% White non-Hispanic, 34% African American non-Hispanic, 24% Hispanic, and 5% from other backgrounds); English proficiency (20% spoke a language other than English); and receipt of welfare cash assistance (36% of primary caregivers were receiving Aid to Families with Dependent Children/Temporary Assistance for Needy Families (AFDC/TANF) at the time of program enrollment).

Program Characteristics

Program implementation. For research purposes, the level of program implementation was defined as the extent to which a program offered services meeting the requirements of the 1995 Early Head Start grant announcement and selected key elements of the revised Head Start Program Performance Standards (USDHHS, 1996). *Full implementation* was defined as substantially implementing, or exceeding expectations for implementing, the key requirements. Rating scales were developed to assess the degree of implementation in key areas and overall. In 1997, we rated 24 areas in (a) early childhood development and health services, (b) family and community partnerships, and (c) program design and management. In 1999, we rated 25 areas.

A rating panel—comprising four national evaluation team members, a representative of the Early Head Start technical assistance network, and another outside expert—used a consensus-based process to assign implementation ratings to each Early Head Start research program, following researchers' site visits to the programs in fall 1997 and fall 1999. (See ACF, 2002b, for a detailed description of the rating scales and process.) The consensus-based ratings were similar to ratings assigned independently by a Head Start Bureau monitoring team member.

The 17 research programs included 6 "early implementers," which became fully implemented by fall 1997 and maintained full implementation in fall 1999; 6 "later implementers," which were not fully implemented in fall 1997 but became so by fall 1999; and 5 "incomplete implementers," which did not achieve full implementation by fall 1999, although they made significant progress in many areas.

Program approaches and services. The programs' approaches to service delivery, as specified in the Head Start performance standards, were chosen by grantees to meet the needs of the particular communities and low-income families they would serve. The 17 research programs included four center-based programs (which provided child development services mainly in center-based child care along with parenting education and a minimum of 2 home visits a year to each family); 7 home-based programs (which provided child development services to families mainly through weekly home visits and at least 2 parent–child group socialization activities a month for each family); and 6 mixed-approach programs (which provided home-based and/or center-based services, either to different families or in combination to families either simultaneously or at different times; ACYF, 1999).

Overall service receipt was comparable across the three program approaches. The mean duration of enrollment among families in center-based programs was 20 months, and children received an average of 1,391 hr of Early Head Start center-based child care. The mean duration of enrollment among families in home-based programs was 22 months, with the average family receiving between two and three home visits per month. The mean duration of enrollment among families in mixed programs was 23 months, and the average family received slightly more than two home visits a month. About 30% of families in mixed programs received Early Head Start center-based care during their program enrollment (these families received an average of 1,400 hr of care). In addition, some families in mixed programs that received home-based services also received child care from community providers who worked in partnership with the Early Head Start program. Home visitors in the home-based programs had the highest levels of education (three fourths had completed at least a two-year college degree), followed by home visitors and teachers in mixed programs (two thirds) and teachers in center-based programs (one third with at least a two-year degree).

Data Collection and Measures

Follow-up data were gathered by data collectors centrally trained and certified as reliable. Data collectors and coders of videotaped parent–child interactions were not informed of families' program status. To obtain data on developmental outcomes, interviewer-assessors attempted to visit families in their homes when the children were 14, 24, and 36 months old. Findings reported here are based on data obtained when children averaged 37.4 months of age.

We report findings for the primary outcomes assessed for child development and parenting. They include measures of cognitive and language development, child social-emotional development, child health, and parenting behavior. Table 2 presents descriptive statistics for each of the measures described and reported (means, standard deviations, internal consistency reliability coefficients, possible range of scores, and sample sizes). The evaluation's technical report includes descriptions of all measures administered (ACF, 2002a).[6]

Child cognitive and language development. Child measures included two direct assessments: (a) the Mental Development Index (MDI) from the Bayley Scales of Infant Development (2nd ed.; Bayley, 1993)—we report the mean MDI and the percentage of children who scored more than one standard deviation below the standardized mean (MDI $<$ 85); and (b) the Peabody Picture Vocabulary Test (3rd ed; PPVT–III; Dunn & Dunn, 1997)

[6] Other measures in the overall study were as follows: Test de Vocabulario en Imagenes Peabody (TVIP), administered to Spanish-speaking children (Dunn, Padilla, Lugo, & Dunn, 1986); child and parent measures from a second, shorter videotaped task in which the child completed two puzzles of different levels of difficulty; the Bayley Behavioral Rating Scale; subscales from the HOME scale; multiple additional self-report measures of parenting and discipline practices, and many reports of services received. (See results section for explanation for selection of outcome measures for this presentation.)

to assess receptive vocabulary—we report the mean PPVT–III standard score and the percentage of children who scored more than one standard deviation below the standardized mean (PPVT–III < 85).

Child social-emotional development. The 19-item aggressive behavior subscale of the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2000), adopted from the Achenbach System of Empirically Based Assessment, provided parent ratings of children's aggressive behavior problems. Observational measures were obtained from videotaped parent–child interactions in a semistructured play task, adapted for this evaluation from the Three Box coding scales used in the National Institute of Child Health and Human Development (NICHD) Study of Early Child Care (NICHD Early Child Care Research Network, 1997). The play task consisted of a 10-min session in which the mother was presented with three cloth bags of toys and invited to play with her child however the child wished; the only request was that she use the three bags in a specified order. Instructions were deliberately vague to elicit naturally occurring behaviors. The sessions were conducted in Spanish or in English according to parental preference and were coded by an observer fluent in the language used. We report on the following child behaviors coded from videotaped parent–child play interactions: (a) sustained attention with objects (the duration and quality of the child's exploring and playing with toys), and (b) engagement of parent (extent to which the child interacts with the parent and communicates positive regard or affect). These behaviors were coded on a 7-point scale ranging from 1 (*very low*) to 7 (*very high*). See discussion of reliability below.

Child health. In the parent interview, we asked parents several questions about their child's health. We report on the following two measures: (a) a global rating of health status over the past year on the 5-point scale (1 = *poor,* 5 = *excellent*) developed for the National Health Interview Survey (Eisen, Ware, Donald, & Brook, 1979) as modified for the IHDP (Gross, Spiker, & Haynes, 1997); and (b) one question from the Parent Services Interview administered 28 months after program enrollment, reporting whether the child had received immunizations since the last interview.

Parenting. We used five parenting measures to assess both supportive and negative aspects of parenting. The total score from the Home Observation for Measurement of the Environment (HOME) measures the quality of stimulation and support available to a child in the home environment (Caldwell & Bradley, 1984; Fuligni, Han, & Brooks-Gunn, 2004). Two measures were coded from the videotaped parent–child semistructured play (described above): (a) parent supportiveness, a composite mean score of sensitivity, positive regard, and cognitive stimulation on a 7-point scale (range: 1 = *very low,* 7 = *very high*); and (b) parent detachment, the degree to which the parent is emotionally unavailable during play, also coded on a 7-point scale (range: 1 = *very low,* 7 = *very high*). The measure *reads to child daily* is a parent self-report of whether the parent reads to the child *every day* or *more than once a day* versus three categories of less often. The measure *spanked child last week* is a parent self-report of physical punishment during the previous week.

Follow-up data collection about service use (including child immunizations reported here) was targeted for 6, 15, and 26 months after random assignment with all sample families (the average data collection times were 7, 16, and 28 months after random assignment). Data were collected by telephone (or in person when necessary) by centrally trained interviewers who used computer-assisted personal interviewing techniques.

Interviewer–Assessor Reliability Certification

Interviewers had to meet certification requirements before they conducted interviews or child assessments with study participants. To be certified to conduct the parent interview, interviewers had to demonstrate that they could administer the interview smoothly, build rapport with the families, and complete the interview accurately. Local research coordinators accompanied interviewers on practice visits and reported to MPR when interviewers met these criteria. To be certified to conduct the Bayley MDI, interviewers had to meet 85% of the certification criteria developed for the study. For the PPVT–III, interviewers had to demonstrate accuracy in completing scoring exercises during training and in administering and scoring the items for two practice administrations. To be certified to conduct the videotaped parent–child interactions, interviewers had to meet 85% of the certification criteria developed for the study. If interviewers did not pass the certification requirements for a specific part of the interview–assessment, they submitted two additional interviews–videotapes for review. After certification, interviewers were periodically checked for adherence to the certification criteria by their local site coordinator and through monitoring by the MPR–Columbia certification team.

Measurement Reliability

We also documented the interrater reliability of the videotape coders. Coding scales were modified from those used in the NICHD Study of Early Child Care (NICHD ECCRN, 1997) and the Newark Observational Study of the Teenage Parent Demonstration program (Spiker et al., 1993) to be appropriate for the Early Head Start population. All coding was done centrally by multiethnic teams of graduate students trained to a criterion level of 85% agreement (exact or within 1 point) on all 7-point scales. After interrater reliability was achieved, intermittent reliability checks were performed on 15% to 20% of each coder's weekly videotape assignments. Coder reliability scores for the 36-month parent and child scales ranged from 89% to 98% agreement within 1 point. Intraclass correlations were computed with a two-way mixed effects model with an absolute agreement definition (McGraw & Wong, 1996) and ranged from .54 to .69 at 36 months, which are considered moderate to substantial (.61 to .80; Landis & Koch, 1977).

To check internal-consistency reliability, we computed Cronbach's coefficient alpha for all outcome measures that yielded summary scores.[7] All except one of the outcome measures reported here had alpha reliabilities of .70 or higher (HOME Support of Language and Learning subscale: α = .67).

Response Rates

There was a 70% response rate for both the 28-month parent services interview and the 36-month parent interview and a 55% response rate for the 36-month direct child assessments. Overall, response rates were similar for program and control groups. To test for attrition bias, we compared the follow-up samples on their baseline characteristics using *t* tests for binary and continuous variables and chi-square tests for categorical variables. We found very few of the distributions of the baseline variables for respondents in the two research groups to be significantly different from each other.[8] Furthermore, none of the joint tests for treatment group differences across all baseline variables considered together was statistically significant (and in the estimation of program impacts, we used regression models to adjust for the few differences in program-control baseline characteristics). Response rates were similar across program approaches, except that response rates for the program group in center-based sites were somewhat higher (ACF, 2002a, chap. II).

[7] We did not compute internal consistency reliability for outcome measures with well-established psychometric properties (such as the Bayley MDI and the PPVT). Published internal consistency reliability data using multiple approaches to computing reliability is considerably greater than .70 (Bayley, 1993; Dunn and Dunn, 1997).

[8] The differences were statistically significant at the 5% level for only 2 of the 48 univariate tests for the 36-month Bayley assessment, only 2 of the tests for the 36-month videotaped assessment, only 2 of the tests for the 36-month parent interview, and only 3 of the tests for the 26-month parent services follow-up interview. These numbers approach the 2 to 3 differences in baseline characteristics that would be expected to be significant by chance.

Table 1
Comparison of the Baseline Characteristics (in Percentages) of All Program and Control Group Members

Variable	Program (N = 1,513)	Control (N = 1,488)	p
Family and primary caregiver characteristics[a]			
Age of mother at birth of focus child			.803
Younger than 20	39.0	39.5	
20 to 25	33.2	32.0	
25 or older	27.9	28.5	
Highest grade completed			.175
Less than 12th grade	47.7	47.8	
12th grade or earned a GED	27.3	29.8	
More than 12th grade	24.9	22.4	
Race and ethnicity			.968
White	37.3	37.1	
Black	34.2	35.0	
Hispanic	23.8	23.4	
Other (Asian or Pacific Islander, American Indian, Eskimo, Aleut)	4.7	4.5	
Primary occupation			.826
Employed	22.9	23.8	
In school or a training program	22.0	21.4	
Neither employed nor in school or training	55.0	54.7	
English language ability			.485
Primary language is English	79.9	78.1	
Primary language is not English, but the applicant speaks English well	9.6	10.3	
Primary language is not English, and the applicant does not speak English well	10.5	11.6	
Living arrangements			.762
Living with a spouse	24.9	25.4	
Living with other adults	38.3	39.1	
Living with no other adults	36.8	35.5	
Adult male present in the household	38.1	39.1	.586
Number of nonfocus children in household under 5 years old			.781
0	64.3	65.1	
1	27.0	26.8	
2+	8.7	8.1	
Number of children in household between 6 and 17 years old			.454
0	64.3	66.4	
1	23.1	21.3	
2+	12.6	12.3	
Number of moves in the past year			.884
0	49.5	49.8	
1	28.9	28.1	
2+	21.6	22.1	
Owns home	11.0	11.1	.907
Household income as a percentage of poverty level			.257
Less than 33	30.2	30.0	
33 to 67	32.5	29.2	
67 to 99	24.0	26.5	
100 or more	13.3	14.3	
Welfare receipt			
AFDC/TANF	35.6	34.7	.627
Food stamps	48.0	47.8	.889
Medicaid	76.6	74.7	.217
SSI	7.0	7.0	.978
WIC	87.5	85.9	.235
Public housing	9.5	8.9	.565
Characteristics of focus child			
Age (in months)			.330
Unborn	24.2	26.5	
Less than 5	36.1	34.7	
5 or more	39.7	38.7	

Table 1 (*continued*)

Variable	Program (N = 1,513)	Control (N = 1,488)	p
Sex			.493
Male	51.7	50.4	
Female	48.3	49.6	
First born	62.3	62.8	.783
Birthweight less than 2,500 g[b]	9.9	8.4	.237
Born more than 3 weeks early[b]	15.8	12.0	.014[a]
Stayed in hospital after birth[b]	18.3	16.0	.178
People concerned about the child's overall health and development[b]	13.0	13.3	.870
Received an evaluation because of concerns about the child's overall health and development or suspected developmental delay[b]	6.0	6.9	.412
Risk categories[c]			
Has established risks[b]	11.6	10.6	.444
Has biological or medical risks[b]	18.3	16.8	.396
Has environmental risks[b]	32.5	36.4	.062[a]
Covered by health insurance[b]	90.1	89.6	.723

Note. Head Start Family Information System application and enrollment forms were completed prior to random assignment. GED = graduate equivalency diploma; AFDC = Aid to Families with Dependent Children; TANF = Temporary Assistance for Needy Families.
[a] The primary caregiver is considered to be an adult regardless of her age. [b] These variables pertain to families with focus children who were born at baseline. [c] These categories of risks are used by most states to identify young children at risk for adverse developmental outcomes. Examples of established risks are a chromosomal abnormality, a congenital birth defect, a sensory impairment, or HIV+/AIDS. Examples of biological or medical risks are congenital heart disease, diabetes, low birthweight, or a severe chronic illness. Examples of environmental risks are parental substance abuse, low maternal education, suspected child abuse or neglect, family social disorganization, or homelessness.
[a]Values sharing subscripts are significantly different from zero at the .10 level, two-tailed test.

Table 2
Descriptive Statistics for Selected Early Head Start Child and Parent Outcome Measures

Outcome	M	SD	Internal consistency reliability	Possible range	N
Child cognitive and language development					
Average Bayley MDI	90.6	12.6	—	49–150	1,658
Proportion with MDI <85	0.30	0.46	—	0–1	1,658
PPVT–III standard score	83.0	15.6	—	40–160	1,424
Proportion with PPVT–III <85	0.52	0.50	—	0–1	1,424
Child social–emotional development					
CBCL Aggressive Behavior	11.1	6.5	0.88	0–38	2,031
Sustained attention with objects during play	4.9	1.0	—	1–7	1,656
Engagement of parent during play	4.7	1.0	—	1–7	1,659
Child health					
Child's health status	4.0	0.98	—	1–5	2,106
Immunizations	0.98	0.13	—	0–1	2,085
Parenting					
HOME total score	27.2	4.8	0.80	0–37	1,807
Supportiveness in play	3.9	0.9	0.82	1–7	1,658
Detachment in play	1.2	0.6	—	1–7	1,659
Reads to child daily (%)	0.54	0.50	—	0–1	2,072
Spanked child last week (%)	0.50	0.50	—	0–1	2,029

Note. Table includes descriptive statistics for the Early Head Start group and the control group, combined. Dashes denote measures with established internal consistency reliability or measures constructed from single items. *M*s and *SD*s were calculated as proportions for percentage variables. Parent interviews, interviewer observations, and assessments of semistructured parent–child interactions were conducted when children were approximately 37 months old. Immunization data are from parent service interviews conducted an average 7, 16, and 28 months after random assignment. MDI = Mental Development Index; PPVT–III = Peabody Picture Vocabulary Test—III; CBCL = Child Behavior Checklist, HOME = Home Observation for Measurement of the Environment.

We observed some differences in the baseline characteristics of respondents and nonrespondents in each research group. Compared with nonrespondents, at the time of enrollment, respondents were better educated and more likely to be employed, to be receiving welfare assistance, and to be married or living with another adult. To address these differences, we constructed sample weights for each instrument as the inverse of predicted response probabilities so that the weighted observable baseline characteristics of respondents were similar to the baseline characteristics of the full sample of respondents and nonrespondents. We constructed sample weights separately for program and control group members to allow for differences in the effects of covariates on response probabilities by research status. The covariates included a wide array of measures from the baseline application and enrollment forms that were correlated with the outcome measures and with response probabilities. Sample weights were constructed separately for outcomes from different data collection instruments and from different waves of data collection. The impact estimates were very similar, regardless of whether these weights were used in the analysis. (Analyses are available from the authors on request, and they can be seen in Appendix D4 of the study's technical report: ACF, 2002a, available at http://www.acf.hhs.gov/programs/opre/ehs/ehs_resrch/index.html#reports and http://www.mathematica-mpr.com/earlycare/ehstoc.asp)

Analytic Approach

The impacts of participation in Early Head Start on child and family outcomes are presented as differences in mean outcomes for the program and control groups. To adjust for any differences in the observable characteristics of program and control group members in the analysis sample resulting from random sampling and interview nonresponse, we estimated regression-adjusted means for each group.[9]

Impacts were obtained as the mean of the regression-adjusted impacts in each site. Sites were given equal weight in the analysis because Early Head Start services are administered at the site level and differ across programs. Two-tailed statistical tests were used to gauge the statistical significance of the estimated impacts; no corrections were made for multiple comparisons. Significant impacts are reported when $p < .05$; impact estimates with $p < .10$ are identified as approaching significance when they contribute to a conceptually consistent pattern of impacts across multiple outcomes. Separate models were estimated for each outcome measure.

To examine the robustness of the analyses reported here, we also estimated impacts under four other assumptions: (a) through the use of a simple differences-in-means approach (in which no explanatory variables are included in the models), (b) by weighting site impacts by their respective sample sizes, (c) by excluding the three sites with the lowest response rates from the analysis, and (d) by both including and excluding weights for nonresponse from the analysis. (Analyses can be found in an appendix to the study's technical report [ACF, 2002a] or are available from the authors on request.)

Although many randomized trials report only "intent-to-treat" impact analyses, an important goal of this study was to produce findings that the federal program office responsible for administering Early Head Start would find credible. Therefore, we estimated program impacts for families that received at least a minimum amount of program services. Because random assignment occurred at the point of eligibility rather than when families began receiving services, a few program group families that enrolled received no services. To estimate impacts for participants only, we divided the impacts for eligible applicants (all who had applied for Early Head Start enrollment and were randomized) by the site's program group participation rate (Angrist, Imbens, & Rubin, 1996; Bloom, 1984).[10] This enabled us to obtain unbiased impact estimates for participants under the assumption that Early Head Start had no effect on nonparticipants. To be confident that this (untestable) assumption holds, we defined program participation conservatively: a *program group family* met the requirements for participation if during a period of 26 months after random assignment it received more than one home visit, met with a case manager more than once, enrolled its child in an Early Head Start center for at least 2 weeks, or participated in a group parent–child socialization activity at least once. Under this definition, 91% of the program families were participants. Participation levels exceeded 90% in 15 sites. Nonparticipants were more likely than participants to be in center-based sites, and they were more likely to be teenage mothers, in school rather than employed, on welfare, and enrolling with firstborn children. Because of the high participation rate, the impacts for all eligible families (the intent-to-treat analysis) are similar to the impacts reported here for participants only (they are available in the

[9] The regression model for each outcome included 16 site indicator variables, 17 interaction terms formed by interacting the treatment status and site indicator variables, and 46 explanatory variables collected at baseline that were selected because they had predictive power in the regression models for key outcome measures and were predictors of interview nonresponse. The categories of explanatory variables reflected families' characteristics and past experiences. They included mothers' age, race–ethnicity, English-language ability, education level, primary occupation, and living arrangements; number of children in the household; poverty level, welfare receipt, and adequacy of resources; previous enrollment in another child development program; mobility during the previous year; age of child at random assignment and at the time of the parent interview or child assessment; low birth weight status; and child's gestational age, gender, and risk categories. Footnote 10 presents the equation that was used for the typical analysis. The impacts were estimated by using a fixed-effects specification rather than a multilevel model because the evaluation sites were selected purposively, not randomly. Thus, the impact estimates pertain only to the 17 sites and do not necessarily generalize more broadly.

[10] This estimator can be derived by expressing the impacts for eligible families (denoted by I) as a weighted average of the impacts for those eligible families that would enroll and receive services if given the chance (I_E) and the impacts for those eligible families that would not enroll (I_{NE}), with weights p and $(1 - p)$, respectively, where p is the program participation rate. If it is assumed that the program has no effects on families that do not participate (that is, if $I_{NE} = 0$), then the impacts per eligible applicant are due entirely to those who enroll in the program, and the impacts per participant can be calculated as the impacts per eligible family divided by p (that is, $I_E = I/p$). This procedure is based fully on the experimental design because it allocates the impacts based on all program and control group members to program participants only. Data on all program and control group members were needed to obtain the impacts for participants. The standard errors of the impacts for participants were adjusted by the estimation error in the program participation rate.

The following system of equations was typically used to estimate impacts by using two-stage least squares (instrumental variable) estimation techniques:

$$S_j{*}P = \delta_j S_j{*}T + u_j \tag{1}$$

$$y = \sum_j \alpha_j (S_j{*}P) + X\beta + \varepsilon, \tag{2}$$

where S_j is an indicator variable equal to 1 if the family is in site j, P is an indicator variable equal to 1 if the program group family participated in Early Head Start (for control group families and program group nonparticipants, $P = 0$), T is an indicator variable equal to 1 if the family is in the program group, y is an outcome variable, X are explanatory variables (that include site indicator variables), ε and the u_js are mean zero disturbance terms, and δ_j, α_j, and β are parameters to be estimated. In this formulation, the estimate of α_j from the second-stage regression represents the impact estimate per participant in site j.

These impacts are average program effects for participants who received different levels of services. Impacts by level of service receipt using propensity score matching procedures are presented elsewhere (ACF, 2002a).

project's final technical report, ACF, 2002a, or from the authors on request).

Program impacts summarized here were estimated for all 17 research sites together and for subgroups of sites defined by program approach and implementation pattern. We obtained subgroup impact estimates by program approach and implementation pattern by comparing the outcomes of program and control group members in those sites. For example, we obtained impacts for center-based programs by averaging the regression-adjusted site-level impacts for the four sites offering center-based services. Thus, the subgroup impacts are based fully on the experimental design. The subgroup impacts must be interpreted carefully, however, because program approaches were not randomly assigned to programs (this would have been inconsistent with Early Head Start guidance that programs should choose their approach based on community needs). Thus, the subgroup impacts describe the effectiveness of specific program approaches for programs that adopted those approaches, given their community contexts and eligible populations.[11] Although the subgroup results cannot be used formally to assess how successful a particular program approach would be if it were adopted more broadly, they provide crucial information that can be generalized to those programs selecting each approach. Such information is useful for future program development.[12]

We conducted tests to gauge the statistical significance of the subgroup impacts, as well as chi-square tests to examine whether differences in the regression-adjusted impacts for participants differed across subgroups. We used these tests, along with the pattern of estimated subgroup impacts and their effect sizes, to interpret the subgroup impact estimates and to identify important variations in impacts across sites with different features. The effect size was calculated by dividing the estimated impact per participant on the outcome measure by the standard deviation of the outcome measure for the control group. This provides a way of comparing impacts across measures in terms of the size of the program–control difference relative to the standard deviation of the measure.

The impact analyses included all sample members for whom 36-month outcome data were available. The Head Start Bureau intended the study results to be applicable to all types of children and families that Early Head Start serves; thus, the analyses included children with disabilities, those who were born with low birth weight, and those who had experienced complications at birth, insofar as assessments and interviews could be completed.

Results

We present results here that highlight the range of domains in which impacts of Early Head Start occurred and indicate how program performance standards and approaches to providing child development services might have contributed to generating those impacts. We have omitted some measures from this presentation because they (a) are redundant with measures included here (e.g., additional measures of discipline and subscales of the HOME); (b) are weaker psychometrically (i.e., child negativity in semistructured play); (c) offer a less direct assessment of a similar outcome (e.g., Bayley behavior ratings compared with measures of child behavior coded from videotapes); (d) constitute a similar rating scale with less ability to predict child outcomes (e.g., behavior during a puzzle challenge task compared with the semistructured play task); (e) focus on non-English-speaking children only (i.e., TVIP), and space precludes full treatment of that segment of our sample; or (f) represent outcomes of less interest developmentally (e.g., parent health and self-sufficiency). The evaluation's technical report includes detailed findings from all measures administered, including several on which the program had no impact and others that had positive impacts (ACF, 2002a).

Overall Impacts of Early Head Start on Children's Development and Parenting

Table 3 presents estimates of impacts of the Early Head Start program. The first two columns show the regression-adjusted means for program group participants and those control group participants who would have been Early Head Start participants if they had been assigned to the program group. The third column shows the estimated impact (program–control difference) per participant and the standard error of the estimate. The fourth column shows the magnitude of the impact in standard deviation units—the effect size.

Children's development. Overall, Early Head Start produced positive impacts on children's cognitive and language development. Early Head Start children scored higher on the Bayley MDI ($M = 91.4$) than did control children ($M = 89.9$), although both groups scored below national norms. The estimated impact of Early Head Start was 1.6 points ($p = .01$). Early Head Start children achieved a higher standardized score on the PPVT–III ($M = 83.3$) than did control children ($M = 81.1$), for an estimated impact of 2.1 points ($p = .02$), and were less likely to score below 85 (51.1% of program group children scored below 85 vs. 57.1% of control group children in that range). The estimated impact of Early Head Start was 6 percentage points ($p = .04$). Effect sizes ranged from .10 to .13 for the cognitive and language development outcomes.

Early Head Start also had positive impacts on several aspects of children's social–emotional development. Early Head Start children were rated by their parents as having lower levels of aggressive behavior ($M = 10.6$; maximum possible score is 38) than were control children ($M = 11.3$); thus, Early Head Start children scored 0.7 points lower, on average ($p = .04$; effect size = .11). In the semistructured play situation, Early Head Start children were rated higher on engagement of their parent and sustained attention to objects, with impacts of 0.2 scale points on each, scored on 7-point scales (p values were .000 and .003, respectively), and effect sizes were .20 (engagement) and .16 (sustained attention).

The health status of Early Head Start and control group children was rated by parents as *very good,* on average, with no difference between the two groups. Immunization rates were very high for both groups, with the Early Head Start group rate marginally higher (1.2%) than the control group rate of 97.7% ($p = .07$).

Parenting behavior. Early Head Start produced positive impacts on emotional support and on parental support for language and learning. Specifically, the average HOME score for control group parents was 27.0 (maximum possible is 37); Early Head

[11] To try to isolate these subgroup effects from other systematic differences in programs or families, we explored the use of hierarchical linear models in which we regressed the 17 site-specific impacts on key program and family characteristics. The subgroup impact estimates produced by using these models were not materially different from those presented. See ACF (2002a, Appendix D) for a detailed discussion of the impact estimates for different program approaches and implementation patterns.

[12] In addition to the subgroup analyses based on program characteristics, the evaluation examined impacts on subgroups of families differing by demographic characteristics measured before random assignment. Space does not permit presentation of those findings here, but positive impacts for children and families were observed in 24 out of the 27 subgroups studied (see ACF, 2002a).

Table 3
Selected Early Head Start Overall Impacts on Children and Parents

Outcome	Program group participants[a]	Control group[b]	Estimated impact per participant[c] (*SE*)	Effect size[d]
Child cognitive and language development				
Average Bayley MDI	91.4	89.9	1.6** (0.63)	.12
Percentage with MDI <85	27.3	32.0	−4.7* (2.43)	−.10
PPVT–III standard score	83.3	81.1	2.1** (0.88)	.13
Percentage with PPVT–III <85	51.1	57.1	6.0** (2.88)	−.12
Child social–emotional development				
CBCL aggressive behavior	10.6	11.3	−0.7** (0.33)	−.11
Sustained attention with objects during play	5.0	4.8	0.2*** (0.05)	.16
Engagement of parent during play	4.8	4.6	0.2*** (0.06)	.20
Child health				
Child's health status	4.0	4.0	−0.0 (0.06)	−.02
Immunizations (%)	99.0	97.7	1.2* (0.67)	.09
Parenting				
HOME total score	27.6	27.0	0.5** (0.22)	.11
Supportiveness in play	4.0	3.9	0.1*** (0.05)	.15
Detachment in play	1.2	1.3	−0.1* (0.04)	−.09
Parent reads to child daily (%)	56.8	52.0	4.9** (2.44)	.10
Spanked child last week (%)	46.7	53.8	−7.1*** (2.49)	−.14
N				
Parent interview	1,083	989	2,072	
Parent services interview	1,075	1,008	2,083	
Parent-child interactions	875	784	1,659	
Bayley MDI	879	779	1,658	
PPVT–III	738	665	1,403	

Note. All impact estimates were calculated by using regression models in which each site was weighted equally. Parent interviews, interviewer observations, and assessments of semistructured parent-child interactions were conducted when children were approximately 37 months old. Immunization data are from parent services interviews conducted an average 7, 16, and 28 months after random assignment. MDI = Mental Development Index; PPVT–III = Peabody Picture Vocabulary Test—III; CBCL = Child Behavior Checklist; HOME = Home Observation for Measurement of the Environment.
[a] A participant is defined as a program group member who received more than one Early Head Start home visit, met with an Early Head Start case manager more than once, received at least 2 weeks of Early Head Start center-based care, and/or participated in Early Head Start group parent-child activities. [b] The control group mean is the mean for the control group members who would have participated in Early Head Start if they had been assigned to the program group instead. This unobserved mean was estimated as the difference between the program group mean for participants and the impact per participant. [c] The estimated impact per participant is measured as the estimated impact per eligible applicant divided by the proportion of program group members who participated in Early Head Start services (which varied by site). The estimated impact per eligible applicant is measured as the difference between the regression-adjusted means for all program and control group members. [d] The effect size was calculated by dividing the estimated impact per participant by the standard deviation of the outcome measure for the control group.
* $p < .10$. ** $p < .05$. *** $p < .01$.

Start parents scored 0.5 points higher ($p = .02$). In the semistructured play situation, Early Head Start parents were rated as more supportive than were control parents (responding to the child's bids for attention, encouraging learning during play, and showing positive regard toward the child). The mean score for Early Head Start parents was 0.1 scale point higher than the control group mean of 3.9 out of a possible 7 ($p = .01$). Early Head Start parents' mean level of detachment during semistructured play (1.2 out of a possible 7) was marginally different from the control group level (difference = 0.1, $p = .09$). A larger percentage of Early Head Start parents (56.8%) than control parents (52.0%) reported reading to their children every day during the previous week (difference = 4.9 percentage points, $p = .047$). Early Head Start parents (46.7%) were less likely than control parents (53.8%) to report that they had spanked the child in the previous week (difference = 7.1 percentage points, $p = .004$). Effect sizes for measures of parenting behavior that showed impacts of Early Head Start ranged from .10 (percentage reading to the child every day) to .15 (rating of supportiveness during semistructured play).

Impacts by Program Implementation Pattern

Children's development. Table 4 summarizes Early Head Start program impacts for subgroups of programs that were early, later, or incomplete in implementing the Head Start standards. The table has three sets of columns, each corresponding to a group of programs defined by their implementation pattern. Within each set are the mean outcomes for the Early Head Start participants, the mean for the control group, the estimated impacts and their standard errors, and the effect sizes. Chi-square tests of whether the difference in impacts across the subgroups was statistically significant were conducted for each outcome, with the results indicated by asterisks on the outcome variable name in the first column.

Table 4
Impacts on Selected Child and Parent Outcomes by Pattern of Program Implementation

Outcome	Early implementers					Later implementers					Incomplete implementers				
	Program group	Control group[a]	Impact estimate per participant[b] (*SE*)		Effect size[c]	Program group	Control group[a]	Impact estimate per participant[b] (*SE*)		Effect size[c]	Program group	Control group[a]	Impact estimate per participant[b] (*SE*)		Effect size[c]
Child cognitive and language development															
Average Bayley MDI	94.1	92.0	2.2*	(1.12)	.17	88.2	86.0	2.2**	(0.98)	.17	92.1	92.1	−0.1	(1.47)	−.00
Percentage with MDI <85***	24.1	27.0	−5.6	(3.76)	−.12	36.5	43.1	−6.6	(4.38)	−.14	23.8	25.8	−2.0	(5.04)	−.04
PPVT–III standard score	86.0	84.8	1.3	(1.44)	.08	78.4	75.2	3.3*	(1.70)	.20	84.8	83.2	1.6	(1.63)	.10
Percentage with PPVT–III <85***	43.1	50.5	−7.5	(4.59)	−.15	65.4	71.2	−5.8	(5.33)	−.12	46.6	51.9	−5.3	(5.96)	−.11
Child social–emotional development															
CBCL aggressive behavior	11.1	11.8	−0.7	(0.62)	−.11	10.8	11.0	−0.2	(0.53)	−.03	9.8	11.6	−1.8***	(0.68)	−.28
Sustained attention with objects during play	5.1	5.0	0.1	(0.09)	.14	4.8	4.7	0.1	(0.09)	.13	5.0	4.8	0.2*	(0.11)	.22
Engagement of parent during play	4.9	4.8	0.1	(0.10)	.11	4.7	4.5	0.2*	(0.09)	.22	4.9	4.5	0.4***	(0.12)	.37
Child health															
Child's health status	4.1	4.1	−0.0	(0.07)	−.03	4.0	3.9	0.1	(0.08)	.10	4.0	4.1	−0.1	(0.10)	−.10
Immunizations (%)	99.1	97.4	1.7	(1.10)	.12	98.3	97.7	0.6	(1.21)	.04	99.5	98.6	0.9	(1.20)	.06
Parenting															
HOME total score	28.3	27.3	1.0**	(0.40)	.20	26.3	26.1	0.2	(0.35)	.04	28.3	27.9	0.5	(0.44)	.09
Supportiveness in play**	4.1	4.1	0.0	(0.08)	.05	3.8	3.6	0.2**	(0.08)	.20	4.1	3.8	0.2**	(0.11)	.23
Detachment in play	1.2	1.2	0.0	(0.04)	.02	1.2	1.3	−0.1*	(0.06)	−.18	1.2	1.2	−0.1	(0.06)	−.10
Reads to child daily***	63.3	52.0	11.3***	(3.97)	.23	49.5	43.3	6.2	(4.10)	.12	57.6	58.9	−1.3	(5.32)	−.03
Spanked child last week***	44.0	52.2	−8.1*	(4.21)	−.16	46.8	55.7	−8.9**	(3.95)	−.18	49.6	56.7	−7.2	(5.55)	−.14
N															
Parent interview	386	358	744			418	362	780			301	281	582		
Parent-child interactions	306	291	597			349	295	644			220	198	418		
Parent services interviews	366	367	733			409	371	780			300	270	570		

Note. All impact estimates were calculated by using regression models in which each site was weighted equally. Parent interviews, interviewer observations, and assessments of semistructured parent-child interactions were conducted when children were approximately 36 months old, and parent services follow-up interviews were conducted approximately 7, 16, and 28 months after random assignment. MDI = Mental Development Index; PPVT–III = Peabody Picture Vocabulary Test–III; CBCL = Child Behavior Checklist; HOME = Home Observation for Measurement of the Environment.
[a] The control group mean is the mean for the control group members who would have participated in Early Head Start if they had been assigned to the program group instead. This unobserved mean was estimated as the difference between the program group mean for participants and the impact per participant. [b] The estimated impact per participant is measured as the estimated impact per eligible applicant divided by the proportion of program group members who participated in Early Head Start services (which varied by site). [c] The effect size was calculated by dividing the estimated impact per participant by the standard deviation of the outcome measure for the control group. [d] Asterisks next to variable names in first column indicate significance levels for chi-square test of differences in impacts across the subgroups.
* $p < .10$. ** $p < .05$. *** $p < .01$.

All program implementation patterns resulted in some favorable impacts on children, but the two groups that fully implemented the performance standards (either early or later) produced more. Programs implementing the standards later had impacts on both cognitive and social–emotional outcomes. Early and later implementers had impacts of 2.2 points on the Bayley MDI, although the impact was significant only for the programs implemented later ($p = .03$). The effect size for both was .17. Early Head Start children in programs implementing the performance standards later showed higher levels of engagement of the parent during semistructured play relative to the control group (difference = 0.2, $p = .02$). Early Head Start children in programs that were incomplete implementers also showed higher levels of engagement of the parent during play (difference = 0.4, $p = .002$) and were rated as less aggressive by parents (difference = 1.8, $p = .01$).

Parenting behavior. Programs that implemented the performance standards early or later had statistically significant impacts on a range of parenting behaviors. Early Head Start parents in programs that implemented early had an average score on the HOME that was 1 point higher than the average score for the control group ($p = .02$), and 63.3% of these parents reported reading to their child daily. This is 11.3 percentage points higher than the percentage of control parents reporting daily reading. Early Head Start parents in programs that implemented later were rated as more supportive during parent–child semistructured play than parents in the control group (difference = 0.2, $p = .02$, effect size = .20). Only one aspect of parenting (supportiveness in play) among Early Head Start parents in programs that were incomplete implementers was significantly different from the control group.

Impacts by Program Approach

Children's development. The second subgroup analyzed is the approach (center-based, home-based, or mixed) that programs adopted for delivering child development and family services. Table 5 shows these impacts. For children in mixed-approach programs, significant impacts occurred in both the language and the social–emotional domains. Early Head Start children in mixed-approach programs scored higher than the control group on the PPVT–III (difference = 3.7, $p = .04$) and showed higher levels of engagement of the parent and sustained attention with objects during semistructured play. Effect sizes for impacts on children in mixed programs ranged from .23 (PPVT–III) to .31 (sustained attention). Early Head Start mixed-approach programs also produced impacts on a range of parenting behaviors.

Parenting behavior. Parents in mixed-approach programs, relative to their control group, were rated as more supportive and less detached during semistructured play, a greater percentage reported reading to the child daily, and a smaller percentage reported spanking in the previous week. Effect sizes for statistically significant impacts on parenting behavior for parents in mixed-approach programs ranged from .21 (supportiveness) to .28 (percentage reading daily).

Early Head Start home-based programs produced impacts in fewer domains than did mixed-approach programs, despite the fact that families in home-based programs constituted the largest program subgroup and thus provided greater power to detect impacts than the other two subgroups. Among early Head Start families in home-based programs, impacts occurred in children's social–emotional development and some aspects of parenting. Specifically, children in home-based programs had higher levels of engagement of the parent during semistructured play than did their control group counterparts. Early Head Start parents in home-based programs were rated as more supportive during semistructured play.

No statistically significant impacts were found among families in center-based programs, but the small size of this subgroup provided less power to detect impacts. In addition, according to the chi-square tests of differences in impacts across subgroups, the impacts for center-based programs did not differ significantly for those in the other program approaches on many of the child and parenting outcomes. This means, for example, that because none of the approaches produced a significant impact on the Bayley MDI, there is no difference in the contribution of each to the overall significant impact of Early Head Start on the MDI.

Impacts by Implementation Pattern Within Program Approach

Within the mixed-approach programs, we estimated impacts for two subgroups with differing implementation patterns—the three programs that were early implementers and the three that were later or incomplete implementers. The early implementers showed a stronger pattern of impacts across several domains of child development and parenting behavior than did the later and incomplete implementers (see Table 6), with effect sizes for statistically significant impacts ranging from .26 to .46.[13]

Discussion

In light of the national Early Head Start program's rapid, large-scale implementation, the significant, favorable, modest-sized impacts of the 17 Early Head Start research programs on a range of child development and parenting outcomes are particularly noteworthy. Guided by the three research questions that this article addresses, we found that (a) overall, Early Head Start programs had significant impacts on a range of child and parent outcomes when the children were 3 years old; (b) programs that were well implemented in relation to the federal program performance standards (whether early or later) produced a greater range of impacts for both children and their parents; (c) impacts were greater for children and parents attending the mixed-approach programs, which combined home- and center-based services, and being fully implemented during the early period increased the number and magnitude of impacts found in mixed-approach programs.

The overall positive impacts included higher performance in children's cognitive and language functioning, as well as a reduction in aggressive behavior as rated by their primary caregiver. The significant reduction in those scoring more than 1 standard deviation below the mean on the PPVT–III (receptive vocabulary), as well as the parallel reduction (approaching significance) in the proportion of children scoring below 85 on the Bayley MDI (cognition), may be important for lessening their risk of needing remedial services at an early age. That Early Head Start can

[13] Because of the small number of center-based programs, we were unable to perform a similar analysis within that subgroup. An analysis of home-based programs by implementation pattern yielded findings that were more variable and more difficult to interpret (results available from the authors on request).

Table 5
Impacts on Selected Child and Family Outcomes at Age 3 by Program Approach

Outcome	Center-based programs				Mixed-approach programs				Home-based programs			
	Program group	Control group[a]	Impact estimate per participant[b] (*SE*)	Effect size[c]	Program group	Control group[a]	Impact estimate per participant[b] (*SE*)	Effect size[c]	Program group	Control group[a]	Impact estimate per participant[b] (*SE*)	Effect size[c]
Child cognitive and language development												
Average Bayley MDI	89.8	88.9	0.9 (1.55)	.07	89.3	87.9	1.4 (1.20)	.11	94.1	92.8	1.2 (0.89)	.10
Percentage with MDI <85***[d]	26.5	36.1	−9.7 (6.11)	−.20	36.1	38.4	−2.2 (4.63)	−.05	20.5	22.0	−1.4 (3.37)	−.03
PPVT–III standard score	83.2	81.8	1.5 (1.91)	.09	82.2	78.5	3.7** (1.74)	.23	84.6	83.1	1.5 (1.35)	.09
Percentage with PPVT–III <85***	52.4	54.7	−2.3 (6.57)	−.05	56.0	67.7	−11.6** (5.20)	−.23	45.6	48.6	−3.0 (4.64)	−.06
Child social–emotional development												
CBCL aggressive behavior	9.6	10.8	−1.2 (0.80)	−.18	10.7	−11.3	−0.6 (0.67)	−.09	11.2	11.7	−0.5 (0.47)	−.08
Sustained attention with objects in play	5.0	5.0	0.0 (0.00)	.01	5.0	4.7	0.3*** (0.10)	.31	5.0	4.9	0.1 (0.07)	.11
Engagement of parent in play	4.9	4.7	0.2 (0.13)	.17	4.7	4.4	0.3*** (0.10)	.30	4.8	4.6	0.2** (0.08)	.19
Child health												
Child's health status	3.9	4.1	−0.2 (0.12)	−.17	4.1	4.1	0.0 (0.08)	.02	4.0	4.0	−0.0 (0.07)	−.04
Immunizations (%)	98.7	98.2	0.5 (2.14)	.03	98.5	97.3	1.2 (1.16)	.08	99.2	98.5	0.7 (0.82)	.05
Parenting												
HOME total score	27.3	26.4	0.9 (0.57)	.19	27.0	26.4	0.6 (0.45)	.11	28.3	28.1	0.2 (0.28)	.04
Supportiveness in play	4.1	4.0	0.1 (0.12)	.09	4.0	3.8	0.2** (0.09)	.21	4.0	3.9	0.1** (0.07)	.16
Detachment in play**	1.2	1.1	0.1 (0.08)	.16	1.2	1.4	−0.2** (0.07)	−.24	1.2	1.3	−0.1 (0.05)	−.09
Reads to child daily***	57.9	50.8	7.0 (6.42)	.14	59.0	45.0	14.0*** (4.47)	.28	54.5	55.7	−1.2 (3.40)	−.02
Spanked child last week***	51.4	61.0	−9.6 (6.38)	−.19	46.6	57.6	−10.9** (4.53)	−.22	44.1	49.6	−5.5 (3.56)	−.11
N												
Parent interview	249	210	459		344	340	693		493	439	932	
Parent-child interactions	227	181	408		252	255	507		396	348	744	
Parent services interview	230	204	434		358	352	710		487	452	939	
Bayley MDI	217	172	389		266	257	523		396	350	746	
PPVT–III	224	168	392		210	229	439		304	268	572	

Note. All impact estimates were calculated by using regression models in which each site was weighted equally. Parent interviews, interviewer observations, and assessments of semistructured parent-child interactions were conducted when children were approximately 36 months old and parent services follow-up interviews were conducted approximately 7, 16, and 28 months after random assignment. MDI = Mental Development Index; PPVT-III = Peabody Picture Vocabulary Test-III; CBCL = Child Behavior Checklist. HOME = Home Observation for Measurement of the Environment.
[a] The control group mean is the mean for the control group members who would have participated in Early Head Start if they had been assigned to the program group instead. This unobserved mean was estimated as the difference between the program group mean for participants and the impact per participant. [b]The estimated impact per participant is measured as the estimated impact per eligible applicant divided by the proportion of program group members who participated in Early Head Start services (which varied by site). [c]The effect size was calculated by dividing the estimated impact per participant by the standard deviation of the outcome measure for the control group. [d]Asterisks next to variable names in first column indicate significance levels for chi-square test of differences in impacts across the subgroups.
** $p < .05$. *** $p < .01$.

Table 6
Impacts on Selected Child and Family Outcomes by Implementation Pattern Within Mixed-Approach Programs

Outcome	Early implementation					Late or incomplete implementation				
	Program group	Control group[a]	Impact estimate per participant[b] (*SE*)		Effect size[c]	Program group	Control group[a]	Impact estimate per participant[b] (*SE*)		Effect size[c]
Child cognitive and language development										
Average Bayley MDI*[d]	93.1	89.5	3.7*	(1.89)	.28	85.4	86.4	−1.0	(1.85)	−.08
Percentage with MDI <85***	27.2	36.4	−9.2	(6.69)	−.20	45.3	43.6	1.7	(7.94)	.04
PPVT–III standard score	85.8	83.4	2.4	(2.39)	.15	78.3	73.4	4.9	(3.22)	.30
Percentage with PPVT–III <85***	45.7	62.6	−16.9**	(7.63)	−.34	66.2	70.2	−4.0	(9.37)	−.08
Child social–emotional development										
CBCL aggressive behavior	11.0	12.0	−1.0	(1.48)	−.15	10.3	10.3	−0.1	(0.81)	−.01
Sustained attention with objects during play	5.1	4.7	0.4***	(0.16)	.42	4.8	4.7	0.2	(0.15)	.17
Engagement of parent during play	4.9	4.5	0.5***	(0.16)	.43	4.6	4.4	0.2	(0.16)	.20
Child health										
Child's health status	4.1	4.2	−0.1	(0.14)	−.14	4.2	4.0	0.2	(0.13)	.17
Immunizations (%)	99.4	97.3	2.2	(1.58)	.15	97.7	98.8	−1.1	(1.95)	−.08
Parenting										
HOME total score	27.8	27.1	0.6	(0.74)	.13	26.3	26.0	0.3	(0.63)	.06
Supportiveness in play	4.1	3.8	0.3	(0.15)	.27	3.8	3.6	0.2	(0.14)	.18
Detachment in play	1.3	1.4	−0.1	(0.11)	−.22	1.1	1.4	−0.2**	(0.10)	−.36
Reads to child daily***	60.4	37.3	23.1***	(6.98)	.46	58.0	50.4	7.6	(6.64)	.15
Spanked child last week***	42.3	55.5	−13.2*	(7.42)	−.26	50.7	61.8	−11.1*	(6.59)	−.22
N										
Parent interview	169	181	350			175	159	334		
Parent-child interactions	122	139	261			130	116	246		
Parent services follow-up interviews	180	195	275			178	157	335		
Bayley MDI	136	153	289			130	104	234		
PPVT–III	125	143	268			85	86	171		

Note. All impact estimates were calculated by using regression models in which each site was weighted equally. Parent interviews, interviewer observations, and assessments of semistructured parent-child interactions were conducted when children were approximately 36 months old, and parent services follow-up interviews were conducted approximately 7, 16, and 28 months after random assignment. MDI = Mental Development Index; PPVT–III = Peabody Picture Vocabulary Test—III; CBCL = Child Behavior Checklist, HOME = Home Observation for Measurement of the Environment.
[a] The control group mean is the mean for the control group members who would have participated in Early Head Start if they had been assigned to the program group instead. This unobserved mean was estimated as the difference between the program group mean for participants and the impact per participant. [b] The estimated impact per participant is measured as the estimated impact per eligible applicant divided by the proportion of program group members who participated in Early Head Start services (which varied by site). [c] The effect size was calculated by dividing the estimated impact per participant by the standard deviation of the outcome measure for the control group. [d] Asterisks next to variable names in first column indicate significance levels for chi-square test of differences in impacts across the subgroups.
* $p < .10$. ** $p < .05$. *** $p < .01$.

influence cognitive and language functioning of young children is consistent with earlier evaluations of early childhood programs for infants and toddlers. Other studies have reported beneficial effects on these outcomes, in some cases showing strong effects on IQ, ranging from one half to two thirds of a standard deviation (e.g., Brooks-Gunn et al., 1993; Campbell & Ramey, 1994; IHDP, 1990; Karoly et al., 1998), with larger effect sizes found in early cohorts when there were few community services (e.g., center-based child care) available to control group families (Wasik et al., 1990). In the case of Early Head Start, larger impacts occurred in the programs that became fully implemented within a year after funding and could tailor their services to families' needs by offering a combination of center- and home-based services. These early implemented mixed-approach programs had stronger impacts, with effect sizes of .28 on Bayley MDI scores and .34 on the percentage of children scoring below 85 on the PPVT-III.

Early Head Start also reduced parent-reported early aggressive behavior, which few early childhood program evaluations have assessed. In the IHDP (1990) study, which did assess aggressive behavior, researchers reported a similar decrease in externalizing behavior, with an effect size comparable to that found for Early Head Start. This decrease is noteworthy because early aggressiveness is predictive of later behavior problems and difficulty with school achievement (Caspi, Moffitt, Newman, & Silva, 1996; Dishion, French, & Patterson, 1995; Tremblay, 2000). Early Head Start programs also produced positive impacts on important aspects of children's social–emotional functioning assessed by direct observation of interactions with the parent—sustained attention to

objects and engagement of the parent during semistructured play. These include elements of children's self-regulation and impulse control that have implications for later learning (Buckner, Mezzacappa, & Beardslee, 2003; Lawson & Ruff, 2004).

The program also produced positive impacts on parent–child interactions, also assessed by direct observation, similar to those found in other early childhood program interventions (for a review, see Brooks-Gunn et al., 2000). As a consequence of their enrollment in the program, Early Head Start parents, compared with control parents, provided home environments that were more supportive of both children's learning and development and the child during play, than did control parents. Program parents were also more likely to read to their child every day. These features of parenting are important predictors of later school achievement (Bradley, Corwyn, Burchinal, McAdoo, & Coll, 2001; NICHD ECCRN, 2002). Early Head Start parents also reported fewer instances of spanking.

Many researchers have called for program evaluations to pay more attention to implementation (Gilliam et al., 2000; Gomby, 1999; McCall & Green, 2004; Wang & Reynolds, 2000). The Early Head Start evaluation contributes to the literature by systematically measuring program implementation and examining its role in producing intervention impacts. The significant child social–emotional impacts for the incompletely implemented programs may reflect emphases in these programs (typically, home visiting and serving teenage parents) on the parent–child relationship and keeping mothers in school.

An important policy question relates to the nature and composition of services that can best meet the needs of low-income families with infants and toddlers. Thus, this evaluation examined the pattern and size of impacts for programs taking center-based, home-based, and mixed approaches. Consistent with the literature, we did not see a significant impact on cognitive and language development in programs that were solely home-based (see Benasich et al., 1992; Sweet & Appelbaum, 2004), but impacts in center-based and mixed-approach programs yielded effect sizes greater than expected (Barnett, 1995; Benasich et al., 1992; Campbell & Ramey, 1994; Karoly, 1998; Ramey & Campbell, 1984; Ramey & Ramey, 1998; Wasik et al., 1990). However, significant impacts were found in social–emotional development in home-based and mixed-approach programs (although more were found in mixed-approach programs). There were significant impacts on several parenting variables in mixed-approach and home-based programs, which is also consistent with the literature, as parenting is typically targeted during home visits (Sweet & Appelbaum, 2004). That there were more and larger impacts in the mixed-approach programs suggests that offering a combination of center-based and home-based services may be a particularly effective way to provide two-generation services. Furthermore, in assessing the effect of implementation within the most effective program approach, we saw that impacts for mixed-approach programs that were fully implemented early were considerably larger than the overall impacts, with effect sizes as large as .46. The impacts in these subgroups suggest that the potential for Early Head Start programs to improve child and family outcomes is greater than the overall impacts indicate.

Noteworthy in the Early Head Start study was the large number of variables across which significant effects were found. No studies have determined the overarching meaning of such a broad pattern of multiple significant impacts, yet the value of impacts on children's development and parenting might be expected to accumulate across outcomes, as would be consistent with the dynamic reciprocity model of human capital in which changes in children and parents each potentially affect the slope of the children's developmental trajectories (Heckman, 2000).

In conclusion, the Early Head Start evaluation tested the impacts of a national infant–toddler program that allows for local community-designed options but has common standards of quality, comprehensiveness, and accountability. We saw that, as Early Head Start programs evolved during the evaluation period, more moved toward having a mixed approach (with both center- and home-based services) as they responded to changing family needs. (In many cases, this move was driven by the new welfare reform legislation enacted in the same month the Early Head Start research programs began enrollment; ACF, 2002b.) Regardless of the program approach programs selected or evolved into, each program had to meet the Head Start Bureau's rigorous performance standards. Just as programs differed in the approaches taken, they varied in their success in meeting these standards. By taking into account both program approach and degree of success in meeting the standards, this evaluation was able to demonstrate that the strongest impacts were achieved by programs that took a mixed approach to services (with both home- and center-based services tailored to family needs) while successfully meeting the performance criteria early in their development.

The significant, favorable, modest-sized impacts of the 17 Early Head Start research programs on a range of child development and parenting outcomes is particularly noteworthy for families living in poverty. The conditions under which the greatest impacts are found provide guidance to other programs that hope to accomplish some of the same goals for their families.

References

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual of ASEBA preschool forms and profiles.* Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families.

Administration for Children and Families. (2002a). *Making a difference in the lives of children and families: The impacts of Early Head Start programs on infants and toddlers and their families.* Washington, DC: U. S. Department of Health and Human Services.

Administration for Children and Families. (2002b). *Pathways to quality and full implementation in Early Head Start Programs.* Washington, DC: U. S. Department of Health and Human Services.

Administration on Children, Youth, and Families. (1999). *Leading the way: Characteristics and early experiences of selected Early Head Start programs. Vol. I: Cross-site perspectives* (Publication U. S. Government Printing Office-2000–519–425/94725). Washington, DC: U. S. Department of Health and Human Services.

Angrist, J., Imbens, F., & Rubin, D. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association, 91,* 444–472.

Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *The Future of Children, 5*(3), 25–50.

Bayley, N. (1993). *Bayley scales of infant development, 2nd ed.: Manual.* New York: The Psychological Corporation.

Benasich, A. A., Brooks-Gunn, J., & Clewell, B. C. (1992). How do mothers benefit from early intervention programs? *Journal of Applied Developmental Psychology, 13,* 311–362.

Bloom, B. (1964). *Stability and change in human characteristics.* New York: Wiley.

Bloom, H. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Review, 8,* 225–246.

Bornstein, M. H. (Ed.). (2002). *Handbook of parenting: Children and parenting.* Mahwah, NJ: Erlbaum.

Bradley, R. H., Corwyn, R. F., Burchinal, M., McAdoo, H. P., & Coll, C. G. (2001). The home environments of children in the United States, Part II: Relations with behavioral development through age thirteen. *Child Development, 72,* 1868–1886.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design.* Cambridge, MA: Harvard University Press.

Brooks-Gunn, J. (2003). Do you believe in magic?: What we can expect from early childhood intervention programs. *Social Policy Report, 17*(1), 1–16.

Brooks-Gunn, J. (2004). Intervention and policy as change agents for young children. In P. L. Chase-Lansdale, K. Kiernan, & R. J. Friedman (Eds.), *Human development across lives and generations: The potential for change* (pp. 293–340). New York: Cambridge University Press.

Brooks-Gunn, J., Berlin, L. J., & Fuligni, A. S. (2000). Early childhood intervention programs: What about the family? In J. P. Shonkoff & S. J. Meisels (Eds.), *Handbook of early childhood intervention* (2nd ed., pp. 549–588). New York: Cambridge University Press.

Brooks-Gunn, J., Klebanov, P. K., Liaw, F. R., & Spiker, D. (1993). Enhancing the development of low birthweight, premature infants: Changes in cognition and behavior over the first three years. *Child Development, 64,* 736–753.

Brooks-Gunn, J., & Markman, L. B. (2005). The contribution of parenting to ethnic and racial gaps in school readiness. *The Future of Children, 15*(1), 139–168.

Buckner, J. C., Mezzacappa, E., & Beardslee, W. R. (2003). Characteristics of resilient youths living in poverty: The role of self-regulatory processes. *Development and Psychopathology, 15,* 139–162.

Caldwell, B. M., & Bradley, R. H. (1984). *Home Observation for Measurement of the Environment: Administration manual, rev. ed.* Unpublished manuscript. Little Rock: University of Arkansas at Little Rock, Center for Applied Studies in Education.

Campbell, F. A., & Ramey, C. T. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child Development, 65,* 684–698.

Caspi, A., Moffitt, T. E., Newman, D. L., & Silva, P. A. (1996). Behavioral observations at age 3 predict adult psychiatric disorders. *Archives of General Psychiatry, 53,* 1033–1039.

Daro, D. A., & Harding, K. A. (1999). Healthy Families America: Using research to enhance practice. *The Future of Children, 9*(1), 152–176.

Dishion, T. J., French, D. C., & Patterson, G. R. (1995). The development and etiology of antisocial behavior. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology* (pp. 421–471). New York: Wiley.

Dokecki, P. R., Hargrave, E. C., & Sandler, H. M. (1983). An overview of the Parent-Child Development Center Social Experiment. In R. Haskins & D. Adams (Eds.), *Parent education and public policy.* Norwood, NJ: Ablex Publishing.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test* (3rd ed). Circle Pines, MN: American Guidance Service.

Dunn, L. M., Padilla, E. R., Lugo, D. E., & Dunn, L. M. (1986). *Examiner's manual for the Test de Vocabulario en Imagenes Peabody (TVIP) Adaptacion Hispanoamericano (Hispanic-American adaptation*). Circle Pines, MN: American Guidance Service.

Eisen, M., Ware, Jr., J. E., Donald, C. A., & Brook, R. H. (1979). Measuring components of children's health status. *Medical Care, 17,* 902–921.

Farran, D. C. (2000). Another decade of intervention for children who are low income or disabled: What do we know now? In J. P. Shonkoff & S. J. Meisels (Eds.), *Handbook of early childhood intervention* (2nd ed., pp. 510–548). New York: Cambridge University Press.

Fuligni, A. S., Han, W., & Brooks-Gunn, J. (2004). The Infant–Toddler HOME in the 2nd and 3rd years of life. *Parenting: Science and Practice, 4,* 139–159.

Garrett, S. G., McKinney, K., Kinukawa, A., Redd, Z., & Moore, K. A. (2003). *Program implementation: What do we know?* Washington, DC: Child Trends.

Gilliam, W. S., Ripple, C. H., Zigler, E. F., & Leiter, V. (2000). Evaluating child and family demonstration initiatives: Lessons from the Comprehensive Child Development Program. *Early Childhood Research Quarterly, 15,* 41–59.

Gomby, D. S. (1999). Understanding evaluations of home visitation programs. *The Future of Children, 9,* 27–43.

Gross, R. T., Spiker, D., & Haynes, C. W. (Eds.). (1997). *Helping low birth weight, premature babies: The Infant Health and Development Program.* Palo Alto, CA: Stanford University Press.

Guralnick, M. J. (1997). *The effectiveness of early intervention.* Baltimore: Brookes.

Heckman, J. (2000). *Invest in the very young.* Chicago: Ounce of Prevention Fund.

Infant Health and Development Program. (1990). Enhancing the outcomes of low-birth weight, premature infants: A multisite, randomized trial. *Journal of the American Medical Association, 263,* 3035–3042.

Johnson, D. L., & Walker, T. (1987). Primary prevention of behavior problems in Mexican-American children. *American Journal of Community Psychology, 15,* 375–385.

Karoly, L. A., Greenwood, P. W., Everingham, S. S., Hoube, J., Kilburn, M. R., Rydell, C. P., et al. (1998). *Investing in our children: What we know and don't know about the cost and benefit of early childhood interventions.* Santa Monica, CA: RAND.

Kitzman, H., Olds, D. L., Sidora, K., Henderson, C. R., Jr., Hanks, C., Cole, R., et al. (2000). Enduring effects of nurse home visitation on maternal life course: A 3-year follow-up of a randomized trial. *Journal of the American Medical Association, 19,* 1983–1989.

Lally, J. R., Mangione, P. L., & Honig, A. S. (1988). The Syracuse University Family Development Research Project: Long-range impact of an early intervention with low-income children and their families. In D. R. Powell (Ed.), *Parent education as early childhood intervention: Emerging directions in theory, research, and practice* (pp. 79–104). Norwood, NJ: Ablex Publishing.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159–174.

Lawson, K. R., & Ruff, H. A. (2004). Early attention and negative emotionality predict later cognitive and behavioural function. *International Journal of Behavioral Development, 28,* 157–165.

McCall, R. B., & Green, B. L. (2004). Beyond the methodological gold standards of behavior research: Considerations for policy and practice. *Social Policy Report, 28*(2), 1–19.

McCall, R. B., Ryan, C. S., & Plemons, B. W. (2003). Some lessons learned on evaluating community-based, two-generation service programs: The case of the Comprehensive Child Development Program (CCDP). *Journal of Applied Developmental Psychology, 24,* 125–141.

McCarton, C., Brooks-Gunn, J., Wallace, I., Bauer, C., Bennet, F., Bernbaum, J., et al. (1997). Results at 8 years of intervention for low birth weight premature infants: The Infant Health and Development Program. *Journal of the American Medical Association, 267,* 2204–2208.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1,* 30–46.

Meisels, S. J., & Shonkoff, J. P. (2000). Early childhood intervention: A continuing evolution. In J. P. Shonkoff & S. J. Meisels (Eds.), *Handbook of early childhood intervention* (2nd ed., pp. 3–31). New York: Cambridge University Press.

Nauta, M. J., & Travers, J. (1982). *The effects of a social program: Executive summary of CFRP's infant-toddler component.* Report submitted to ACYF, Office of Human Development Services, U. S. Department of Health and Human Services. Cambridge, MA: Abt Associates.

NICHD Early Child Care Research Network. (1997). The effects of infant child care on infant-mother attachment security: Results of the NICHD Study of Early Child Care. *Child Development, 68,* 860–879.

NICHD Early Child Care Research Network. (2002). Early child care and

children's development prior to school entry: Results from the NICHD Study of Early Child Care. *American Educational Research Journal, 39,* 133–164.

Olds, D. L., Henderson, C. R., Kitzman, H. J., Eckenrode, J. J., Cole, R. E., & Tattlebaum, R. C. (1999). Prenatal and infancy home visitation by nurses: Recent findings. *The Future of Children, 9*(1), 44–65.

Ramey, C. T., & Campbell, F. A. (1984). Preventive education for high-risk children: Cognitive consequences of the Carolina Abecedarian Project. *American Journal of Mental Deficiency, 88,* 515–523.

Ramey, C. T., & Ramey, S. L. (1998). Early intervention and early experience. *American Psychologist, 53,* 109–120.

Rutter, M. (2000). Resilience reconsidered: Conceptual considerations, empirical findings, and policy implications. In J. P. Shonkoff & S. J. Meisels (Eds.), *Handbook of early childhood intervention* (2nd ed., pp. 651–682). New York: Cambridge University Press.

Shonkoff, J. P., & Phillips, D. A. (Eds.). (2000). *From neurons to neighborhoods: The science of early child development.* Washington, DC: National Academy Press.

Spiker, D., Ferguson, J., & Brooks-Gunn, J. (1993). Enhancing maternal interactive behavior and child social competence in low birth weight, premature infants. *Child Development, 64,* 754–768.

St. Pierre, R. G., & Layzer, J. I. (1999). Using home visits for multiple purposes: The Comprehensive Child Development Program. *The Future of Children, 9,* 134–152.

St. Pierre, R. G., Layzer, J. I., Goodson, B. D., & Bernstein, L. (1997). *National impact evaluation of the Comprehensive Child Development Program: Final report.* Cambridge, MA: Abt Associates.

Sweet, M. A., & Appelbaum, M. I. (2004). Is home visiting an effective strategy? A meta-analytic review of home visiting programs for families with young children. *Child Development, 75,* 1435–1456.

Tremblay, R. E. (2000). The development of aggressive behavior during childhood: What have we learned in the past century? *International Journal of Behavioral Development, 24,* 129–141.

U. S. Department of Health and Human Services, Administration for Children and Families. (1995). Early Head Start program grant availability: Notice. *Federal Register, 60,* 14548–14578.

U. S. Department of Health and Human Services, Administration for Children and Families. (1996). Head Start program: Final rule. *Federal Register, 61,* 215.

Wang, M., & Reynolds, A. J. (2000). Early childhood learning: Programs for a new age. *The National Center on Education in the Inner Cities Review, 3,* 1–2.

Wasik, B., Ramey, C. T., Bryant, D. M., & Sparling, J. J. (1990). A longitudinal study of two early intervention strategies: Project CARE. *Child Development, 61,* 1682–1696.

Yoshikawa, H. (1995). Long-term effects of early childhood programs on social outcomes and delinquency. *The Future of Children, 5*(3), 51–75.

Received August 19, 2004
Revision received March 5, 2005
Accepted March 29, 2005 ■

Instructions to Authors

For Instructions to Authors, please consult the March 2005 issue of the volume or visit www.apa.org/journals/dev and click on the "Instructions to authors" link in the Journal Info box on the right.