1

W. S. Barnett

Long-Term Effects on Cognitive Development and School Success

Persistent Effects on Cognitive Development and School Success

The contribution of early care and education (ECE) to the cognitive development and school success of children who are economically and socially disadvantaged has become a vital public issue. Experts generally agree that ECE programs can produce short-term gains in disadvantaged children's performance on standardized tests of

Adapted from Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *The Future of Children* (Winter) 5(3):25–50, by permission of Center for the Future of Children, The David and Lucile Packard Foundation.

intelligence and academic ability and that some preschool programs have reduced later grade retention and special-education placement. However, there is a great deal of disagreement about the true nature of these effects, whether they persist, what other effects might be produced, and what is required to produce meaningful long-term effects (e.g., Zigler & Muenchow, 1992; Locurto, 1991; Seitz, 1990; Barnett, 1992; Schweinhart, Barnes, Weikart, Barnett, & Epstein, 1993; Consortium for Longitudinal Studies, 1983; Haskins, 1989; Spitz, 1986, 1991; Datta, 1983). These disagreements are so severe that they make it difficult to develop research-based ECE policy recommendations, and uncertainty and skepticism regarding long-term effects undermines support for existing public programs like Head Start.

This chapter seeks to resolve at least some of the disagreements and provide a stronger knowledge base for policy making through a critical review of research on the effects of ECE on children from low-income families. Results of the review are used to address the following questions: (1) Can ECE programs produce meaningful long-term effects on the cognitive development and school success of economically disadvantaged children? (2) How do the effects of large-scale government programs such as Head Start and public school preschool education compare to those of model programs? (3) Do persistent effects require or benefit from continuation of intervention beyond the preschool years?

Research on Short-Term Effects

Hundreds of studies have examined the immediate and short-term effects of ECE programs. These studies are found in two largely separate streams of research: one on the effects of ordinary child care on children from all backgrounds and the other on the effects of ECE interventions specially designed to improve the cognitive development of economically disadvantaged children. These two research streams have tended to be conducted from different perspectives. Initially, research on child care focused on possible negative effects on the mother-child relationship and social development with distinctly less attention to possible effects on cognitive development. More recently, child-care research has become more nuanced and begun to examine the effects on cognitive development of variations in the quality of care and possible interactions with the child's home environment and family circumstances. Research on interventions for disadvantaged children

initially emphasized the potential for positive effects on cognitive development, especially as measured by IQ. Over time, interest grew in the effects of interventions on other aspects of cognitive development, on school success, and on socialization.

Child-care research presents no consistent evidence that child care per se is harmful to child development regardless of the age at which a child begins out-of-home care (Lamb & Sternberg, 1990; Zaslow, 1991). Variations in the quality of child care appear to be important determinants of the impact of child care, though the effects of quality seem to vary with child and family characteristics. Higher-quality child care has been found to be associated with better cognitive and social development contemporaneously and into the first few years of school (Lamb & Sternberg, 1990; Phillips, McCartney, & Scarr, 1987; Zaslow, 1991; Helburn & Culkin, 1995). A new large-scale study of infant child care conducted by the NICHD Early Child Care Research Network (1996) found evidence that the frequency of insecure attachment was increased by poor quality care, increased hours of care, and instability of care arrangements, *but only when* the child's mother was rated as "insensitive" in her interactions with the child.

Another recent investigation found that age at entry or years of experience and type of care (child's own home, other home, or center) during the preschool years influenced the reading and math achievement of children at ages five and six (Caughy, DiPietro, & Strobino, 1994). Effects were positive for children from impoverished homes. For these children, earlier entry/more years produced a larger effect on reading scores, and center-based care had a larger effect on math scores. Conversely, effects were negative for children in the highest income families. The interaction appears to be due to differences in the quality of home environments rather than to income per se: children whose home environments were relatively poor (as measured by Caldwell's HOME) gained the most, while children whose home environments were very highly supportive of cognitive development and socialization had lower scores if they had been in other care. This study of effects on achievement is especially interesting because two other studies failed to find positive effects of ordinary child care on IQ (the PPVT) for four-year-old children (Baydar & Brooks-Gunn, 1991; Desai, Chase-Lansdale, & Michael, 1989).

From the intervention research, it appears that programs designed for disadvantaged children, including large-scale public programs, can produce immediate effect sizes for IQ and achievement of about 0.5 standard deviations, equivalent to about 8 IQ

points (White & Casto, 1985; McKey et al., 1985; Ramey, Bryant, & Suarez, 1985). Somewhat smaller average effect sizes were found for immediate effects on socioemotional outcomes such as self-esteem, academic motivation, and social behavior. On average, these estimated effects declined over time and were negligible several years after children exit the programs. However, some programs produced sizeable gains that persisted into the school years for IQ, achievement, grade retention, and special-education placement. A variety of different approaches produced positive effects, but the magnitude of initial effects appears to be roughly related to a program's intensity, breadth, and amount of involvement with children and their families (Ramey, Bryant, & Suarez, 1985).

One difficulty in interpreting the research results is that most studies relied on natural variation in participation in ECE, making it difficult to separate the effects of ECE from the effects of family characteristics that influence ECE enrollment decisions (e.g., parents' income, socioeconomic status, education, and attitudes toward education and child rearing). Experiments in which children are randomly assigned to specially designed ECE programs are extremely valuable because they make it easier to separate program effects from family background effects. Thus, short-term results from relatively new randomized experiments are worth summarizing briefly.

The CARE study randomly assigned children (N=57) to three conditions: a high-quality, full-day, year-round ECE program and home visits for parent education from shortly after birth to age five, home visits alone, and a control group (Roberts et al., 1989; Wasik et al., 1990). At age five, the ECE plus home visits group had higher IQs and better language skills (but did no better on a reading test) than the other two groups despite substantial participation of the others in community child care. No effects were found on parenting.

The Infant Health and Development Program (IDHP) study is an eight-site randomized trial (N=985) of ECE from birth to age three for low-birth-weight infants primarily, but not entirely, from low-income families (IHDP, 1990; Brooks-Gunn et al., 1994). The program consisted of weekly home visits for the first year directed at both parents' and children's needs, followed by biweekly home visits and full-day ECE for the child from age one to age three. By age three, the program had increased children's IQs by thirteen points and improved child behavior slightly as measured by the Child Behavior Checklist. Small effects were found on maternal employment (one month more), but not on maternal education or fertility. Effects on the children were no longer apparent for the full sample at age five and age eight follow-up. However, modest effects on IQ

and achievement remained at ages five and eight for children above 2,000 grams birth weight (McCarton et al., 1997).

The Comprehensive Child Development Program (CCDP) sought to increase and improve social, health, and education services for families with young children through home visitors who served as case managers (St. Pierre and Lopez, 1994). Families (N=4,411, 21 sites) had a child under one year of age at study entry and were to be visited twice monthly. The CCDP substantially increased mother's participation in parenting education, mental-health services, and education, slightly increased children's use of health services, and substantially increased children's participation in formal ECE. After several years, only extremely small effects were found for mothers and children.

Even Start provided parenting education, adult education, and early-childhood education plus a variety of supporting services (St. Pierre et al., 1993). Extent and duration of services varied widely across families, though nearly all children received some early-childhood education. Experiments at five sites with three- and four-year-old children (N=164) found small positive effects on a measure of school readiness skills two and one half years after program entry. Small effects also were found on one aspect of home environment (reading materials) and on parents' expectations for children's academic success.

In sum, ECE has important impacts on cognitive development and abilities associated with school success immediately and in the short term. Effects appear to depend on program quality and the child's home environment and are larger for well-designed, intensive ECE interventions than for ordinary child care or programs focused on increasing families' use of existing health, social, and educational services. Some, but not all, studies report that effects decline after children leave the ECE program.

Research on Long-Term Effects

Studies were selected for review of research on long-term effects if they met four criteria: (1) children entered the program before age five, except for Head Start, which mostly serves three and four year olds but serves five year olds under some circumstances, such as when public kindergarten is not available; (2) the program served economically disadvantaged children; (3) at least one measure of cognitive development, school progress, or socialization was collected at or beyond age eight (third grade); and (4) the research design

provided a no-treatment comparison to a comparable group or adjusted for socioeconomic differences. These criteria excluded studies of children who were not economically disadvantaged, studies of kindergarten, case studies of individual children, simple before-and-after comparisons of children in ECE, and studies that compared disadvantaged children to more advantaged children without any statistical adjustments for differences in family background. The requirement for follow-up to at least third grade allowed sufficient time to observe fade-out in effects (Caldwell, 1987). Thirty-eight studies were identified that met the review criteria. This is a larger number of long-term studies than in previous reviews including the well-known quantitative syntheses (White & Casto, 1985; McKey et al., 1985).

Program and Study Characteristics

The thirty-eight studies were divided into two categories for review based on the nature of the ECE program and the research design. In fifteen studies, researchers developed their own ECE programs to study the effects of exemplary programs. Some of these model programs might be characterized as family-support programs in today's terminology; most, if not all, of them worked with parents in some way. In twenty-three other studies, researchers investigated the effects of ongoing, large-scale, public ECE programs; twelve studied Head Start programs, seven examined public-school programs, and four studied a mix of Head Start and public-school programs.

Model-Program Studies

The fifteen studies of model programs are described in Table 1-1. Generally, the model ECE programs are likely to have been of higher quality than the large-scale public programs. Reasons for this include: (1) the close supervision and direction of experts, (2) highly qualified staff, and (3) low child-staff ratios and small group size. These advantages were made possible by higher levels of funding per child than are available to Head Start and public-school programs. In all but one study the majority of children were African American. The Houston Parent Child Development Center (PCDC) served Hispanic American families. The average level of mother's education was under twelve years in all studies, and under ten years in five studies. Three model-program studies limited their target populations in ways that could have affected their results. The Harlem Training Project served only boys. The Perry Preschool

study selected children based on low IQ scores, and its sample had substantially lower IQs at age three than children in other studies. The Milwaukee study selected children whose mothers had low IQs (below 75).

As can be seen from Table 1-1, the model programs varied in entrance age, duration, services provided, and historical context (1962 to 1980). Most of the comparison children began formal education at kindergarten, but, especially in the later studies, it is likely that significant percentages of the comparison groups attended a preschool or child-care program (as this became more common and Head Start and public school preschool programs grew). For example, in the Abecedarian study, which enrolled newborns between 1972 and 1980, two thirds of the control group attended an ECE program for twelve months or more by age five (Burchinal, Lee, & Ramey, 1989). Clearly, this could lead to some underestimation of the effects of ECE programs.

Head Start and Public School Programs

The twenty-three studies of Head Start and public school ECE programs are identified and described in Table 1-2. None of these programs took children before age three, and most served children part-day for one school year at age four. Class size and child-teacher ratio tended to be higher than in model programs. Head Start programs had broader missions than most public school programs; their goals included improving health and nutrition, and providing services to parents and the community (Zigler & Styfco, 1993). The programs studied seem generally representative of public programs for poor children over the past several decades. In three studies, ECE program participation was associated with differential school-age programs. In the Cincinnati Title I study most full-day kindergarten students had attended preschool and most half-day kindergarten students had not. In the two Child Parent Center (CPC) studies, services began in preschool and continued as enriched education through third grade.

Research Design

Three key aspects of research design are described in Tables 1-1 and 1-2 for model-program and Head Start and public school program studies. These are: (1) the ways in which the comparison groups were formed, (2) initial and follow-up sample sizes, and (3) length of follow-up. Each of these has important implications for the validity and interpretation of study findings.

Table 1-1.
Model Early Childhood Programs

Program Name (Years of Operation) (Sources)	Program Description	Ages of Participation	Research Design/ Methodological Concerns
1. Carolina Abecedarian (1972–1985) (Campbell & Ramey, 1993, 1994; Campbell, 1994)	Preschool-age: full-day child care School-age: parent program	Entry: 6 weeks to 3 months Exit: 5 to 8 years	Randomized.
2. Houston Parent Child Development Center (1970–1980) (Andrews et al., 1982; Johnson & Walker, 1991)	Home visits Full-day child care Center-based program for parents	Entry: 1 to 3 years Exit: 3 to 5 years	Randomized High attrition.[c]
3. Florida Parent Education Project (1966–1970) (Jester & Guinagh, 1983)	Home visits Twice weekly part-day preschool (ages 2 to 3 years)	Entry: 3 to 24 months Exit: 3 years	Initially randomized with one group, and additional control group members added at 24 months. Not randomized.[d] High attrition. School-administered tests.[e]
4. Milwaukee Project (1968–1978) (Garber, 1988)	Full-day child care Job and academic training for mothers	Entry: 3 to 6 months Exit: 3 years	Groups of 3 to 4 children assigned alternately to E and C groups. Small sample.
5. Syracuse Family Research Program (1969–1975) (Lally, Mangione, & Honig, 1988)	Home visits Full-day child care	Entry: 6 months Exit: 5 years	Matched comparison group selected at 36 months. Not randomized.
6. Yale Child Welfare Research Program (1968–1974) (Seitz, Rosenbaum, & Apfel, 1985; Seitz & Apfel, 1994)	Home visits Full-day child care Pediatric care Developmental screenings	Entry: Prenatal Exit: 30 months	Two comparison groups for same neighborhoods for first follow-up. Matched comparison group selected from follow-up at 30 months. Not randomized. School-administered tests.
7. Verbal Interaction Project (1967–1972) (Levenstein, O'Hara, & Madden, 1983)	Home visits	Entry: 2 to 3 years Exit: 4 years	Six groups with three matched comparison groups. Not randomized.

Initial Sample Size[a]	Follow-up Sample Size	Time of Follow-up	IQ[a,b]	School Outcomes[a]
E = 57 C = 54	Age 8 E = 48 C = 42 Age 15 E = 48 C = 44	8, 12, and 15 years	Age 12; E > C E = 93.7 C = 88.4	Achievement test: E > C at age 15 Special education: E < C at age 15: E = 24%, C = 48% Grade retention: E < C at age 15: E = 39%, C = 59%
E = 97 C = 119	School date E = 50 C = 87 IQ data E = 39 C = 78	Grades 2 to 5	Not measured	Achievement tests: E = C, but positive trend Grades: E = C Bilingual education: E < C E = 16%, c = 36% Special education: E = C, grades 2 to 5 E = 27%, C = 31% Grade retention: E = C, grades 2 to 5 E = 16%, C = 29%
E = 288 C = 109	E = 83 C = 24	Grades 4 to 7	E = C (grades 4 to 7) E = 83.1 C = 79.8	Math achievement: E > C Reading achievement: E = C Special education: E < C, grade 7 E = 23%, C = 54% Grade retention: E = C, grade 7 E = 28%, C = 29%
E = 20 C = 20	E = 17 C = 24	Grade 4 Grade 8	Grade 8: E > C E = 101 C = 91	Achievement tests: E = C, but positive trend Grades: E = C Special education: E = C, grade 4 E = 41%, C = 89% Grade retention: E = C, grade 4 E = 29%, C = 56%
E = 82 C = 72	Parents E = 52 C = 42 Children E = 49 C = 39	Grades 7 to 8	E = C, age 5 on Stanford-Binet	Teacher ratings: E > C, but for girls only Grades: E > C, but for girls only Attendance: E > C, but for girls only
E = 18 C = 18	Age 7 to 8 E = 17 C1 = 33 C2 = 31 Age 10 E = 16 C = 16	Age 7 to 8 and age 10	E = C at age 10	Achievement tests: E = C Attendance: E > C Teacher ratings: E = C, but positive trend for boys only Special education: E = C E = 25%, C = 50%
E = 111 C = 51	E = 79 C = 49	Grades 3	E>C at grade 3 E = 101.9 C = 93.6	Achievement test: E > C Special education: E < C, grade 7 E = 14%, C = 39% Grade retention: E = C, grade 7 E = 13%, C = 19%

Table 1-1. (*continued*)
Model Early Childhood Programs

Program Name (Years of Operation) (Sources)	Program Description	Ages of Participation	Research Design/ Methodological Concerns
8. Early Training Project (1962–1967) (Gray, Ramsey, & Klaus, 1982, 1983)	Home visits Summer part-day preschool program	Entry: 4 to 5 years Exit: 6 years	Randomized. School-administered tests.
9. Experimental Variation of Head Start (1968–1969) (Karnes, Schwedel, & Williams, 1983)	Part-day preschool program	Entry: 4 years Exit: 5 years	Post hoc comparison group from same communities. Not randomized. High attrition. School-administered tests.
10. Halem Training Project (1966–1967) (Palmer, 1983)	One-to-one tutoring or child-directed play	Entry: 2 to 3 years Exit: 4 years	Comparison group recruited from children born 1 to 2 months later.
11. High/Scope Perry Preschool Project (1962–1967) (Weikart, Bond, & McNeil, 1978; Schweinhart et al., 1993; Barnett, Young, & Schweinhart, this volume)	Home visits Part-day preschool program	Entry: 3 to 4 years Exit: 5 years	Randomized.
12. Howard University Project (1964–1966) (Herzog, Newcomb, & Cisin, 1974)	Part-day preschool program	Entry: 3 years Exit: 5 years	Comparison group from neighboring tracts. Not randomized.
13. Institute for Developmental Studies (1963–1967) (Deutsch, Deutsch, Jordan, & Grallo, 1983)	Home visits Part-day preschool program Parent center school (K–3)	Entry: 3 years Exit: 9 years	Randomized. High attrition. School-administered tests.
14. Philadelphia Project (1963–1964) (Beller, 1983)	Home visits Part-day preschool program	Entry: 4 years Exit: 5 years	Matched comparison group from same kindergarten classes. Not randomized. School-administered tests.
15. Curriculum Comparison Study (1965–1967) (Miller & Bizzell, 1983, 1984)	Part-day preschool program Kindergarten program	Entry: 4 years Exit: 5 or 6 years	Post hoc comparison group from original pool. Not randomized. School-administered tests.

a. Throughout Table 1-1, E refers to the experimental or intervention group, and C refers to the control or comparison group. Outcomes listed as E>C or E<C were statistically significant at the $p<.05$ level.
b. IQs were measured using the WISC or WISC-R, unless otherwise noted.
c. Results may be biased because of high attrition rates.

Initial Sample Size[a]	Follow-up Sample Size	Time of Follow-up	IQ[a,b]	School Outcomes[a]
E=44 C=21	E=36 C=16	Post-high school	E=C age 17 E=78.7 C=76.4	Achievement tests: E=C Special education: E<C, grade 12 E=5%, C=29% Grade retention: E=C E=58%, C=61% High school graduation: E=C E=68%, C=52%
E=116 C=24	E=102 C=19	Post-high school	E<C at age 13 E=85.0 C=91.0	Achievement tests: E=C, but positive trend Special education: E=C, grade 7 E=13%, C=15% Grade retention: E=C, grade 7 E=10%, C=16%
E=244 C=68	E=168 C=51	Grade 7	E=C at age 12 E=92.1 C=88.9	Math achievement: E>C Reading achievement: E<C Grade retention: E<C, grade 7 E=30%, C=52%
E=58 C=65	E=58 C=65	Post-high school	E=C at age 14 E=81.0 C=81.0	Achievement tests: E>C Grades: E>C Special education: E=C, grade 12 E=37%, C=50% Grade retention: E=C, grade 12 E=15%, C=20% High school graduation: E>C E=67%, C=49%
E=38 C=69	E=30 C=69	Grade 4	Not measured	Grade retention: E=C E=33%, C=47%
E=312 C=191	E=63 C=34	Grade 7	Not measured	Special education: E=C E=0%, C=13% Grade retention: E=C E=23%, C=43%
E=60 C=53	E=44 C=37	Post-high school	E>C at age 10 on Stanford-Binet E=98.4 C=91.7	Achievement test: E=C, but positive trend Special education: E=C, grade 12 E=5%, C=6% Grade retention: E=C, grade 12 E=38%, C=53%
E=244 C=68	E=168 C=51	Post-high school	Not measured	Special education: E=C, grade 12 E=32%, C=63% Grade retention: E=C, grade 12 E=26%, C=58% High school graduation: E=C E=67%, C=53%

d. Results may be biased because children were not randomly assigned to experimental and control or comparison groups.

e. Results may be biased because school-administered tests were used to measure achievement.

Table 1-2.
Large-Scale Public Early Childhood Programs[a]

Program Name (Years of Operation) (Source)	Ages of Participation	Design	Initial Sample Size[b]
1. Child-Parent Center (1965–1977) (Fuerst & Fuerst, 1993)	Entry: 3 or 4 years Exit: 9 years	Compared former CPC children with non-CPC children from same feeder schools.	E=684 C=304
2. Child-Parent Center II (1983–1985) (Reynolds, 1994a, 1994b, 1993)	Entry: 4 or 5 years Exit: 6 years	Compared former CPC children with several other groups.	Unknown
3. Cincinnati Title I Preschool (1969–1970; 1970–1971) (Nieman & Gastright, 1981)	Entry: 4 or 5 years Exit: 6 years	Compared children who attended full-day kindergarten and mostly had preschool with children who attended half-day kindergarten and mostly had no preschool.	E=688 C=524
4. Maryland Extended Elementary Pre-K (1977–1980) (Eckroade, Salehi, & Carter, 1988; Eckroade, Salehi, & Wode, 1991)	Entry: 4 years Exit: 5 years	Compared attenders to nonattenders, including only children continuously enrolled in school district (kindergarten to grade 5).	Unknown
5. New York State Experimental Prekindergarten (1975–1976) (State Ed. Dept., Univ. of the State of NY, 1982)	Entry: 3 or 4 years Exit: 5 years	Compared attenders with children in same district on waiting list and with children in other districts with no prekindergarten program.	1800[h]
6. Florida Prekindergarten Early Intervention Cohort 1 (King, Cappelini, & Gravens, 1995)	Entry: 4 years Exit: 5 years	Compared Pre-K early-intervention children with children from same schools who qualified for free/reduced lunch.	Unknown
7. Florida Prekindergarten Early Intervention Cohort 2 (King, Cappelini, & Rohani, 1995)	Entry: 4 years Exit: 5 years	Compared Pre-K early-intervention children with children from same schools who qualified for free/reduced lunch.	Unknown
8. Florida Chapter I (King, Rohani, & Cappelini, 1995)	Entry: 4 years Exit: 5 years	Compared children screened into with those screened out of Chapter I Pre-K based on a test (DIAL-R)	E=103 C=121
9. Detroit Head Start and Title I Preschool (1972–1973) (Clark, 1979)	Entry: 4 years Exit: 5 years	Compared children who had attended Head Start or Title I preschool with children who were eligible but did not attend.	Unknown
10. D.C. Public Schools and Head Start (1986–1987) (Marcon, 1990, 1993)	Entry: 4 years Exit: 5 years	Compared children who had attended public school or Head Start with children in same kindergartens who had not.	E=372 C=89
11. Philadelphia School District Get Set and Head Start (1969–1970; 1970–1971) (Copple, Cline, & Smith, 1987)	Entry: 4 years Exit: 5 years	Compared children in enriched K-3 program (follow-through) who had and had not attended preschool.	E=1,082 C=1,615
12. Seattle DISTAR and Head Start (1970–1971) (Evans, 1985)	Entry: 4 years Exit: 5 years	Compared children who had attended Head Start and DISTAR with matched children from same school and grades.	E=92 C=unknown

Follow-up Sample Size	Time of Last Follow-up	School Outcomes[b]	Methodological Concerns
E=513 C=244	Post-high school	Achievement test: E>C at grade 2, E=C at grade 8 High school graduation: E>C E=62%, C=49%	Not randomized.[c] No pretest.[d] School-administered tests.[e]
E=757 C=130	Grade 7	Achievement tests: E>C for grades K to 7 Special education: E<C, E=12%, C=22% Grade retention: E<C, E=24%, C=34%	Not randomized. No pretest. School-administered tests.
E=410 C=141	Grade 8	Achievement tests: E>C for grades 1, 5, 8 Special education: E=C, grade 8 E=5%, C=11% Grade retention: E=C, grade 8 E=9%, C=12%	Not randomized. No pretest. School-administered tests.
E=356 C=306	Grade 8	Achievement tests: E>C for grades 3, 5, 8 Special education: E<C, grade 8 E=15%, C=22% Grade retention: E<C, grade 8 E=31%, C=45%	Not randomized. No pretest. High attrition.[f] School-administered tests.
E=1,348 C=258	Grade 3	Achievement tests: E>C in kindergarten E=C in grade 1 Special education: E=C, E=2%, C=5% Grade retention: E<C, E=16%, C=21%	Not randomized. High attrition.
E=350 C=352	Grades 3 and 4	Achievement tests: E>C in kindergarten E=C in grades 1 to 3, E<C in grade 4 Special education: E=C, E=25%, C=25% Grade retention: E=C, E=3%, C=3% Disciplined: E<C, E=11%, C=32%	Not randomized. No pretest. High attrition. School-administered tests. Pre-K E children attended schools in poorer communities. First year of program operation.
E=983 C=1,054	Grades 3 and 4	Achievement tests: E>C in kindergarten E=C in grades 1 to 4 Special education: E=C, E=17%, C=15% Grade retention: E<C, E=9%, C=13%	Not randomized. No pretest. High attrition. School-administered tests.
E=54 C=65	Grade 8	Achievement tests: E>C in grades 1,2,4,7,8 E=C in grades 5,6 (no data for grade 3)	Not randomized. High attrition. School-administered tests.
Unknown	Grade 4	Achievement tests: E>C in grade 4	Not randomized. No pretest. School-administered tests. Bias toward no effect.[g]
E varies C varies	Grades 4 and 5	Achievement tests: E=C in grades 3 to 5 Special education: E=C, grade 4 E=10%, C=9% Grade retention: E=C, grade 4 E=31%, C=38%	Not randomized. Bias toward no effect. High attrition.
E=688 C=524	Grades 4 to 8, varies by cohort	Achievement test: E=C Grade retention: E<C	Not randomized. No pretest. Bias toward no effect. High attrition. School-administered tests.
E=44 C=20	Grades 6 and 8	Achievement tests: E=C, but positive trend in grades 6 and 8	Not randomized. No pretest. High attrition. School-administered tests.

Table 1-2. (*continued*)

Large-Scale Public Early Childhood Programs[a]

Program Name (Years of Operation) (Source)	Ages of Participation	Design	Initial Sample Size[b]
13. Cincinnati Head Start (1968–1969) (O'Piela, 1976)	Entry: 4 years Exit: 5 years	Compared third graders who had attended Head Start with those who had not.	Unknown
14. Detroit Head Start (1969–1970) (Pinkleton, 1976)	Entry: 4 years Exit: 5 or 6 years	Compared children who had attended Head Start with children in Title I elementary programs.	Unknown
15. ETS Longitudinal Study of Head Start (1969–1970; 1970–1971) (Shipman, 1970, 1976; Lee et al., 1990)	Entry: 4 or 5 years Exit: 5 or 6 years	Compared children who had attended Head Start with siblings who had not, using fixed-effects model and percentile scores.	1,875
16, Hartford Head Start (1965–1966) (Goodstein, 1975)	Entry: 4 years Exit: 5 years	Compared children who had attended Head Start with low-income children who had not.	293
17. Kanawha County, West Virginia Head Start[a] (1973–1974) (Kanawha Bd. of Ed., 1978)	Entry: 4 years Exit: 5 years	Compared children who had attended Head Start with low-income children who had not.	Unknown
18. Montgomery County, Maryland Head Start (1970–1971; 1974–1975; 1978–1979) (Hebbeler, 1985)	Entry: 4 years Exit: 5 years	Compared children who had attended eight or nine months with those who had attended one month or less.	E=1,915 C=619
19. NBER-NLSCM Head Start (1979–1989) (Currie & Thomas, 1995)	Entry: 3 to 5 years Exit: 5 to 6 years	Compared children who had attended Head Start with low-income children who had not, using a fixed-effects model and raw scores.	6,676
20. New Haven Head Start (1968–1969) (Abelson, 1974; Abelson, Zigler, & DeBlasi, 1974)	Entry: 4 years Exit: 5 years	Compared children who had attended Head Start with those who had not.	E=61 C=48
21. Pennsylvania Head Start (1986–1987) (Reedy, 1991)	Entry: 3 to 5 years Exit: 5 to 6 years	Compared children who had attended Head Start with children who had applied but had not been admitted.	E=98 C=unknown
22. Rome, Georgia Head Start (1966) (McDonald & Monroe, 1981)	Entry: 3 to 5 years Exit: 5 to 6 years	Compared children who attended Head Start with all children in first grade in disadvantaged schools in 1966.	E=130 C=88
23. Westinghouse National Evaluation of Head Start (1965–1966) (Westinghouse Learning Corp. & Ohio University, 1969)	Entry: 4 or 5 years Exit: 5 or 6 years	Compared children who attended Head Start with those who did not (matched within grade).	Unknown

a. Programs are grouped such that public-school program studies are listed first, followed by program studies involving both public-school programs and Head Start, and then all Head Start studies.

b. Throughout Table1-2, E refers to the experiemental or intervention group, and C refers to the control or comparison group. Outcomes listed as E>C or E<C were statistically significant at the p<.05 level.

c. Results may be biased because children were not randomly assigned to experimental and control or comparison groups.

Follow-up Sample Size	Time of Last Follow-up	School Outcomes[b]	Methodological Concerns
Unknown	Grade 3	Achievement tests: E=C in grade 3	Not randomized.[c] No pretest.[d] Bias toward no effect.[g]
Unknown	Grade 4	Achievement tests: E>C in grade 4	Not randomized. No pretest. School-administered tests.[e] Bias toward no effect.
852	Grade 3	Achievement tests: E>C in grade 1, E=C in grades 2 and 3	Not randomized. High attrition.[f] Bias toward no effect.
E=148 C=50	Grade 6	Achievement tests: E=C in grade 6 Special education: E=C, E=5%, C=10% Grade retention: E<C, E=10%, C=22%	Not randomized. No pretest. High attrition. School-administered tests.
Unknown	Grade 3	Achievement tests: E=C in grade 3	Not randomized. No pretest. High attrition. School-administered tests.
E=186 C=112	Grade 11	Achievement tests: E=C, but negative trend in most grades, C>C in grade 11	Not randomized. No pretest. High attrition. Fixed-effects model assumes no family effects. Used percentile scores that have floor effect for blacks.
762	Grade varies up to 12	Achievement tests: E>C, whites only Grade retention: E<C, whites only	Not randomized. No pretest. High attrition. Fixed-effects model assumes no family effects. Used percentile scores that have floor effect for blacks.
E=35 C=26	Grade 3	Achievement tests: E>C in grade 1, E=C in grade 3 Grade retention: E<C, E=18%, C=35%	Not randomized. No pretest. High attrition. Bias toward no effect.
E=54 C=18	Grade 3	Achievement tests: E=C, but positive trend in grades 2 and 3	Not randomized. No pretest.
E=94 C=60	Post-high school	Achievement tests: E>C in grade 5, E=C in grades 6 and above Special education: E<C, E=11%, C=25% Grade retention: E=C, E=51%, C=63% High school graduation: E>C, E=50%, C=33%	Not randomized. No pretest. School-administered tests.
E=1,988 C=1,992	Grades 1 to 3	Achievement tests: E>C in grade 1, E=C in grades 2 and 3	Not randomized. No pretest. Bias toward no effect.

d. No pretest was given to assess/control for initial differences between groups.
e. Result may be biased because school-administered tests were used to measure achievement.
f. Results may be biased because of high attrition rates.
g. Design flaws bias the estimated effect of the program on children's achievement toward zero.
h. The numbers of children in experimental and comparison groups were not reported separately.

Model-Program Study Designs

Seven model-program studies formed comparison groups by randomly assigning children to experimental and control groups from the same pool of potential participants or by using procedures that approximated random assignment (field studies rarely carry off anything perfectly).[1] This increases confidence that estimated effects in these studies are due to the program rather than to preexisting (though perhaps unmeasured) differences between program and comparison groups. However, the benefits of random assignment can be lost as the result of severe attrition (loss of study participants over time) or small sample size, and small sample size can severely limit the power of a study to detect important effects. Only two of these experimental studies began with sample sizes larger than thirty in each group and had low attrition throughout follow-up—the Abecedarian and Perry Preschool studies. Two other experimental studies (Milwaukee and the Early Training Project) began with extremely small sample sizes, which rendered random assignment less useful and provided these studies with very little power to detect even fairly large effects. The remaining four experimental studies suffered massive attrition that could have invalidated the initial random assignment.[2]

The other eight model-program studies constructed comparison groups, usually at a later date. Some of the approaches to constructing comparison groups seem likely to have created group differences that favored the ECE group. The two curriculum studies formed no-ECE groups after the fact by selecting children who had not attended another ECE program. This eliminated from the potential comparison pool those children whose parents sought out early educational experiences for them and were most comparable to the treatment group families with respect to parental attitudes and behavior regarding education. In the Harlem Training Project, attrition during a waiting period prior to entry at age three may have introduced differences favoring this later entry group as it had a higher IQ prior to treatment than the control group (Lazar et al., 1982). The Yale Child Welfare Research Program study obtained a control group thirty months after it selected the program group using the same clinic records used to identify the program group. However, the program group was invited to receive child care and other services while the comparison group was invited to participate in data collection. There was sufficient rejection of the offer in both cases to significantly influence group composition. Moreover, the passage of time before

the comparison group was contacted meant that many who moved were lost from the comparison group. Those moving without leaving a forwarding address might have been the least economically and socially successful. Finally, three times the number of months of clinic records were required to obtain the program group as the comparison group suggesting that the population using the clinic had changed or that methods used to select the two groups differed (Seitz, Rosenbaum, & Apfel, 1985).

Head Start and Public School Research Designs

All of the large-scale, public program studies used quasi-experimental designs. Some constructed comparison groups from waiting lists or other groups of children thought to be similar to program children. Others simply relied on natural variation in program attendance within a sample. Both strategies raise questions about the comparability of the groups due to self-selection and administrative selection. Self-selection occurs when some parents exert more effort to obtain educational opportunities for their children including ECE programs, good neighborhood schools groups, good teachers within schools, and good educational experiences outside of school. The educational success of their children is unlikely to be comparable even without the benefit of preschool. The results of administrative selection are less clear. Programs might seek to enroll the most needy, those easiest to recruit, or those thought most likely to gain from the program. The Head Start and public school program studies are at a distinct disadvantage in dealing with this problem compared to the model-program studies. Not only did they not use random assignment, but because comparison groups were not identified prospectively, there are no pretest measures of children's cognitive abilities to offer as evidence that the groups were initially the same or to use to adjust later measures.[3]

Some of the Head Start and public school program studies used statistical adjustments for variations in family background characteristics to try to eliminate possible biases introduced by differences between program and comparison groups. Several of the most recent studies employed complex statistical procedures that explicitly attempt to remove the effects of selection. However, the extent to which these statistical methods produce more accurate estimates of effects is unclear, and alternative approaches can produce conflicting results (Barnett & Camilli, 1997; Campbell, 1991; Cook, 1991).

Findings of Long-Term Studies

Most long-term research has focused on the effects of ECE on cognitive ability and school success. Results in these domains are reported in Table 1-1 for model programs and Table 1-2 for large-scale programs. Outcome measures included are IQ, achievement, grade retention, special-education placement, and high school graduation. After a review of these findings, results are summarized for the relatively small number of studies reporting results for socialization and parent outcomes.

IQ Effects

All of the model-program studies found that their ECE programs produced IQ gains at some point. In most cases IQ effects were sustained until school entry at age five, at which time ten studies reported effects between 4 and 11 IQ points, the Milwaukee study reported a gain of 25 points, and the Syracuse study reported no effect. Three studies did not measure IQ at school entry. Data on the persistence of IQ effects is provided by Table 1-1 which reports the most distant follow-up comparison of IQs. The two experimental studies that enrolled infants in full-day educational child-care programs reported the largest initial effects (Milwaukee & Abecedarian) and found that some IQ gain persisted at least into adolescence.[4] The other studies that enrolled infants did not find persistent effects, but both were quasi-experimental and one ceased serving children before age three.

None of the large-scale program studies provided IQ data on Stanford-Binet or WISC IQ tests comparable to the data from the model-program studies. A small number of studies provided results on the Peabody Picture Vocabulary Test (PPVT), and the WLC study administered the Illinois Test of Psycholinguistic Abilities (ITPA). Whether these tests should be considered comparable to IQ tests is questionable. With one exception, no effects are found on these measures after school entry.

The exception is a study by Currie and Thomas (1995), which finds that Head Start produced persistent effects on PPVT scores for white but not African American children using data from the National Longitudinal Survey. Their results were challenged by Barnett and Camilli (1997) who found serious limitations in the data and methods employed by Currie and Thomas. Among the most important of these: their methodology relies on questionable assumptions and leads to the selection of a biased sample; the PPVT-R percentile scores analyzed by Currie and Thomas exhibit a

number of problems including an extremely strong floor effect that does not lift with age for African Americans; and alternative analyses that are equally plausible produce different results and raise questions about the meaningfulness of any analyses employing the NLS data.

Achievement Effects

Five of eleven model-program studies with achievement test data found statistically significant positive effects beyond grade three. Evidence of effects was strongest in the studies that randomized assignment to program and control groups. The Abecedarian and Perry Preschool studies found achievement effects persisting to ages fourteen and fifteen. The Florida Parent Education study found effects through grade four. The Milwaukee study found that effects were statistically significant only to grade two.[5] The ETP and IDS programs did not find effects on achievement. In contrast to the experimental studies, none of the quasi-experimental model-program studies found persistent effects on achievement, though some found statistically significant initial effects.

The achievement test results of the Head Start and public school studies were as variable as the results of the model-program studies. Of the twenty-four studies reporting achievement-test results, nine found significant positive effects at latest follow-up. The other fifteen studies found no effects or that the effects faded out. Fade-out occurred early in most cases, but was not found until at least sixth grade in three studies. Note that studies finding no effect do not necessarily imply that there were no effects before third grade as several measured achievement only at grade three or later.

A naive interpretation of the results of these studies would be to say that most ECE programs have failed to produce long-lasting gains in achievement for disadvantaged children. Based on this conclusion, one might seek to identify the characteristics of successful programs or to find explanations for fade-out in subsequent school experiences (e.g., the poor schools attended by disadvantaged children especially in the inner cities). However, this conclusion is incorrect, and the subsequent search for sources of fade-out is premature. Instead, the evidence of fade-out in achievement appears to result largely from flaws in research design and very high attrition rates for achievement-test data, which reduced sample size (thereby decreasing the statistical power to detect effects) and biased estimated effects toward zero. Studies that found no effects or fade-out were vulnerable to selective attrition because

they obtained their achievement test data from schools' routine testing programs or they suffered from another design flaw that produced a similar problem even though they administered their own achievement tests.

The most common source of achievement test data in these studies was standardized tests routinely administered by schools. Although this strategy provided data at low cost, it had several unfortunate consequences. First, the quality and uniformity of test administration can be expected to be lower when testing is done for entire classes by teachers rather than individually by well-trained testing specialists. Second, some data are lost simply because the tests used vary from school to school and year to year. Third, schools' testing programs administer tests by grade so that children who are retained in grade are not tested with their age cohort. Many studies collected the data by grade and simply lost data on children who had been retained. Even in the rare cases where data on children behind grade level are added later, the scores are not comparable because they are obtained at different ages. Fourth, children expected to perform poorly are systematically excluded from school testing. The use of routine testing to hold schools accountable places pressure on school administrators to remove poor performers from the test pool at each grade level (McGill-Franzen & Allington, 1993). Many schools do not test children in special-education classes. Poor students are more frequently absent and are more likely to miss tests (sometimes because they have been encouraged to miss them).

Studies relying on school-administered tests at best have test scores with lower reliability and smaller sample sizes, both of which would reduce their ability to detect program effects. At worst, they systematically lose the more poorly performing students from year to year as the cumulative percentage of children retained in grade, placed in special education, or otherwise omitted from testing grows. The result is that any differences between program and comparison groups are gradually "erased" as grade level rises and the children for whom achievement tests are available become more similar across the two groups.

Some studies had idiosyncratic flaws that led to similar biases in achievement-test data, even though tests were specially administered for the studies. For example, the New Haven Head Start study individually administered achievement tests, but only to children at expected grade level. As there was significantly less grade retention in the program group over time, this had the effect of gradually equating the tested program and control groups on