The benefits of Head Start are under increased scrutiny as Congress debates its reauthorization. How effective is Head Start, and how can it be improved? We provide a current overview and critical evaluation of Head Start research and discuss implications of this research with an eye toward informing debate. There has been a good deal of controversy over whether Head Start produces lasting benefits, dating back to its early years. Our review finds mixed, but generally positive, evidence regarding Head Start’s long-term benefits. Although studies typically find that increases in IQ fade out over time, many other studies also find decreases in grade retention and special education placements. Sustained increases in school achievement are sometimes found, but in other cases flawed research methods produce results that mimic fade-out. In recent years, the federal government has funded large-scale evaluations of Head Start and Early Head Start. Results from the Early Head Start evaluation are particularly informative, as study participants were randomly assigned to either the Early Head Start group or a control group. Early Head Start demonstrated modest improvements in children’s development and parent beliefs and behavior. The ongoing National Head Start Impact Study, which is also using random assignment, should yield additional insight into Head Start’s effectiveness. We conclude with suggestions for future research.

**Key words:** early education, Head Start, long-term benefits, policy

HEAD START is our nation’s foremost federally funded provider of educational services to young children in poverty. Since 1965, more than 21 million children have participated in this comprehensive child development program (U.S. Department of Health and Human Services, 2003a). As a comprehensive program, in addition to its educational services, Head Start also provides social, health, and nutritional services to children and their low-income parents. When Early Head Start was established in 1994, the program was expanded to serve even younger children (from birth to age 3) and their families. By 2002, the Head Start program reported funding more than 910,000 children with a budget of $6.5 billion (U.S. Department of Health and Human Services, 2003a). However, Head Start remains a promise unfulfilled. Nearly 10 years after Congress authorized full-funding, Head Start’s budget is still insufficient to serve all eligible children or deliver uniformly high-quality services to all enrolled.

As one of the most prominent educational and social programs in the United States, Head Start has attracted both proponents and detractors. How effective is Head Start as an early education program for disadvantaged children? What are the long-term benefits associated with participation in Head Start? These are questions that are reconsidered in each authorization cycle, when the program’s benefits come under increased public scrutiny. Head Start was most recently reauthorized by Congress in 1998 and was scheduled for reauthorization again in 2003, although this process has not yet been completed. This article critically reviews the research on Head Start and other early education programs for at-risk children. We also discuss the implications of this research for issues that are likely to arise during reauthorization. Finally, we present recommendations for future studies of Head Start.
THE WESTINGHOUSE STUDY

Controversy over the benefits of Head Start dates back to its earliest years, when a study by Westinghouse Learning Corporation and Ohio University (1969) reported that the program produced few sustained effects. This was the first prominent effort to investigate Head Start’s impacts over time. Former Head Start children identified in first, second, and third grades were compared to schoolmates within the same grades who had not participated in Head Start, with a focus on cognitive and social-emotional development. Children from the Head Start and comparison groups were matched within grades on other important characteristics including ethnic group, gender, socioeconomic status (SES), and kindergarten attendance.

The Westinghouse study was immediately and widely criticized on methodological grounds (Condry, 1983). However, no one appears to have noticed at the time the most serious methodological flaw. The post hoc selection of the 2 groups literally equated the children on grade level. This biases between-group comparisons to the extent that differences in grade retention rates and special education placements truncated the samples, thereby eliminating the higher percentage of lower performing children from the comparison groups. The most obvious evidence that the comparison group does not represent comparable cohorts is that an increasing age gap is found moving across the grades with children in the third-grade comparison group significantly older than the third-grade Head Start children. Despite this and other evidence of methodological problems, the Westinghouse study continues to be cited in policy debates as evidence that Head Start does not produce sustained educational benefits for children in poverty.

FINDINGS FROM SHORT- AND LONG-TERM STUDIES OF HEAD START

Since the publication of the Westinghouse study, Head Start has continued to draw researchers’ attention. A number of longitudinal studies have followed former participants over time to gather more information about the benefits associated with Head Start. This research can be divided into 2 general categories: short-term and long-term studies. For the purposes of this review, we consider studies with immediate outcome measures and longitudinal studies with outcome measures taken earlier than third grade to be short-term studies, and consider studies with outcomes measured in third grade or later to be long-term studies.

A brief summary of the key findings from the short-term studies follows, since our primary interest is in Head Start’s long-term benefits. Evidence of short-term benefits of preschool programs including Head Start has been thoroughly reviewed elsewhere (Barnett, 2004; McKey et al., 1985; Nelson, Westhues, & MacLeod, 2003; Ramey, Bryant, & Suarez, 1985; White & Casto, 1985). Studies have generally shown that programs for children at risk, including Head Start, result in increases of 0.5 standard deviations in IQ and achievement. Estimated impacts on measures of social behavior, self-esteem, and academic motivation typically are slightly smaller.

A recent short-term study by Abbott-Shim, Lambert, and McCarty (2003) is particularly notable for using random assignment of eligible 4-year-olds who had applied to a large Head Start program with a waiting list. This procedure allowed the researchers to rule out selection bias as an influence on results. Abbott-Shim et al. (2003) found that Head Start participants benefited substantially compared to nonparticipants in the areas of receptive vocabulary and phonemic awareness and had more positive health-related outcomes, for example, they were more likely to be current on their immunizations. And the parents of Head Start children reported more positive health and safety habits than the parents whose children did not attend Head Start. Because of the strength of the research design used in this study, these outcomes provide strong support for the short-term effectiveness of Head Start.
Some past reviewers (Haskins, 2004; McKee et al., 1985; White & Casto, 1985) have found that positive impacts of Head Start and early childhood programs for disadvantaged children decrease over time and eventually fade altogether. However, recent meta-analyses of longitudinal studies (Gorey, 2001; Nelson et al., 2003) suggest that effects persist over time although there may be some diminution of effects over the long term. These findings are consistent with the work done by Barnett, Young, and Schweinhart (1998), who used causal modeling to show that long-term effects of early childhood education are built upon short-term effects.

Reviews focused on long-term studies of early education programs serving economically disadvantaged children (e.g., Barnett, 1998, 2004) find that the evidence regarding Head Start’s long-term outcomes is mixed. In a recent examination of Head Start’s long-term cognitive effects, Barnett (2004) identified only 39 studies in which educational programs included treatment and control groups, served children from low-income families, began during or before the preschool years, and were followed up with cognitive or academic measures at least through third grade, of which 15 were studies of “model” programs and 24 were studies of large-scale public programs. Twelve of the public program studies focused on Head Start, and an additional 4 included both Head Start and public school programs. Several of the model program studies, but none of the large-scale public program studies, employed random assignment.

Studies of model programs typically show initial gains in children’s IQ scores that fade out over time (Barnett, 2004). Studies of large-scale programs have less often measured IQ, although the Peabody Picture Vocabulary Test has sometimes been used as a proxy for verbal IQ, making it more difficult to evaluate whether Head Start produces persistent IQ gains. However, it is likely that initial increases in IQ scores by Head Start children also fade out over time. Findings regarding other types of benefits are more promising. Studies of both model and large-scale programs find achievement effects. In some studies the effects persist, in others effects on achievement cease to be statistically significant over time. Fade-out is frequently associated with high attrition over time or with other design flaws that affect the collection of achievement test data. Yet, decreases in children’s later rates of grade retention and special education placements are found in most studies of model and large-scale programs. This apparent inconsistency often can be explained by differences in data collection procedures that lead to greater, more-biased attrition for achievement test data (Barnett, 2004). Few studies have measured impacts on high school graduation, but those with the largest samples reported statistically significant positive impacts (Barnett, 1998).

Overall, it appears that model programs and large-scale programs such as Head Start have similar types of effects, but the studies of model programs found effects of greater magnitude (Barnett, 2004). Given the variation in populations, programs, and contexts across studies, it is difficult to identify a single cause for this difference in effectiveness. Yet, it seems highly plausible that programs such as Head Start lack the type of funding necessary to produce the levels of intensity and quality achieved in better funded model programs with the direct result that they are less effective. Several studies provide direct evidence in support of this argument (Barnett, 1998). Some of this is discussed below in the context of findings from key studies of non-Head Start preschool interventions.

OUTCOMES FROM OTHER PRESCHOOL INTERVENTION PROJECTS

The Carolina Abecedarian Project (Campbell & Ramey, 1994, 1995) is one of the most notable studies of a model program to provide high-quality early education services to at-risk children. Participants were identified in the 1970s as infants, on the basis of their parents’ low-income status as well as other risk factors predictive of cognitive difficulties
in childhood. The sample ($N = 111$), which was primarily African American, was divided into experimental and control groups by random assignment. Experimental group children attended the full-day, year-round Abecedarian program until age 5. Another randomization took place before children started school, with half the members of both the control and the experimental groups receiving an additional 3-year intervention. Thus, participants in this study received from 0 to 8 years of intervention services, with variation in its timing. Follow-up results have now been reported through age 21 (Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001; Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002).

Findings from the Abecedarian Project show that the program produced large initial effects that persisted long after the intervention ended (Campbell et al., 2001, 2002; Campbell & Ramey, 1994, 1995). At the age 21 follow-up, Campbell et al. (2002) found that program effects were strongest for young adults who had taken part in the (5-year) preschool phase of the intervention. When compared to the preschool control group, these adults showed stronger performance on measures of academic skills and IQ. At age 21, they also were more likely to be enrolled in 4-year colleges, were better educated overall, and were more likely to hold skilled employment. Further, cost-benefit analysis of the Abecedarian Project (Masse & Barnett, 2002) shows that its overall benefits outweigh its costs, on the order of $4 saved for every dollar spent on the preschool intervention (present value discounted at a real rate of 3%).

Research on the Chicago Child-Parent Centers (CPC; Reynolds, Temple, Robertson, & Mann, 2002) provides evidence of the long-term effects of a public-school-operated preschool program. The CPC program began in 1967 and classrooms are located in or near public schools in Chicago’s highest poverty neighborhoods. From age 3 until age 5, participants attend 2.5 hour classes 5 days a week during the school year and a 6-week summer program is also generally provided. After attending kindergarten, participants receive less intensive services through the public schools until age 9. Longitudinal follow-ups of the CPC cohort born in 1980 have been completed through age 21, on the basis of 2 study groups created beginning in 1985: former participants in the preschool and kindergarten phases of the CPC program ($N = 989$) and a comparison group of nonparticipants ($N = 550$). Members of the comparison group were matched to former preschool participants using SES and other demographic factors. Reynolds and colleagues (2002) report positive long-term outcomes from CPC across a wide range of domains. These include persistent gains in reading achievement (age 14), lower rates of grade retention and special education, lower rates of reported child maltreatment (ages 4–17), lower rates of juvenile arrests, and higher rates of educational attainment. A cost-benefit analysis estimates that the CPC preschool program yields an economic return far exceeding its cost (Reynolds et al., 2002).

**RECENT RESEARCH ON HEAD START’S LONG-TERM OUTCOMES**

Although long-term longitudinal evaluations of benefits associated with the Head Start program have been rare, several recent studies have sought new evidence. In a follow-up to the Head Start Planned Variation study conducted from 1969 to 1972, Oden, Schweinhart, Weikart, Marcus, and Xie (2000) compare 22-year-olds who attended Head Start at age 4 to others who had not attended, in 2 communities, 1 in Florida ($N = 424$) and 1 in Colorado ($N = 198$). Former Head Start participants were located as young adults, and a post hoc comparison group was constructed using young adults who had lived on the same streets or in the same high-poverty neighborhoods (Census tracts) as the Head Start participants. Members of the comparison group had not attended Head Start or any other early education program. However, perhaps because many children from the communities’ lowest SES families had attended
Head Start, the Head Start group was slightly lower in SES than the non–Head Start comparison group. Statistical adjustments for these and other differences were made in the data analysis, to facilitate the process of drawing meaningful conclusions from comparisons between the 2 groups.

Few statistically significant differences were found between the Head Start and non–Head Start comparison groups (Oden et al., 2000). However, the direction and pattern of results suggests possible long-term benefits. Most notably, in the Florida sample, girls who had attended Head Start were significantly more likely to graduate high school or earn a GED (95% vs 81%) and significantly less likely to have been arrested at age 22 (5% vs 15%) than were girls in the non–Head Start comparison group.

The lack of strong results in this study may be due to methodological limitations that led to difficulties in adequately controlling for preexisting differences between the Head Start group and the non–Head Start group. All of the initial advantages of the comparison group may not be captured by the difference in socioeconomic status, and the statistical adjustments cannot be guaranteed to produce an unbiased estimate of the impact of Head Start on the more disadvantaged participant group. Oden and colleagues recommend that more rigorously designed studies be developed to obtain stronger evidence.

Janet Currie and colleagues have employed creative statistical approaches to estimate the long-term effects of Head Start from national data sets with self-reported Head Start participation (Currie & Thomas, 1995, 1999; Garces, Thomas, & Currie, 2000). These studies estimate within family differences among siblings where one child is reported to have attended Head Start and another not. One strength of these studies is that they employ data collected across the nation. Limitations include error in self-reported participation and highly restrictive assumptions about the reasons for, and the consequences of, differences in Head Start participation among siblings (Barnett & Camilli, 2002; Currie, 2001). Most of the limitations seem likely to lead to an underestimation of long-term benefits (Currie, 2001). For example, they assume that Head Start has no benefit for siblings and that parents engage in no compensating behaviors to generate more equal outcomes among siblings. These assumptions are unlikely to be true and thus bias downward the estimated effects from comparing siblings (Barnett & Camilli, 2002).

Currie and colleagues find long-term effects for subpopulations: higher long-term Peabody Picture Vocabulary Test scores, less grade repetition, and more high school graduation and college attendance for whites and Latinos and fewer criminal charges and convictions for African Americans. This variation by ethnicity is not predicted a priori, and the lack of persistent academic improvements for African Americans is inconsistent with the findings of randomized trials of other interventions with African American samples. Barnett and Camilli (2002) conducted similar analyses with one data set employed by Currie and Thomas (1995) and found no persistent gains for either white non-Latino or African American children. They caution that limitations of the data and the potential for substantive violations of the analytical assumptions are so serious that such estimates of Head Start’s long-term effects should not be relied upon for public policy purposes.

RECENT FEDERALLY SPONSORED EVALUATIONS OF HEAD START

The federal government has renewed its emphasis on funding large-scale scientific evaluations of Head Start and Early Head Start in recent years. These national, longitudinal studies have been sponsored by the Administration on Children, Youth, and Families to provide more details about the services provided by these programs, as well as better information regarding the progress made by participants and their families.

The Family and Child Experiences Survey (FACES; Zill et al., 2001) of Head Start children
and their families was the first of these studies, beginning in 1997. Of primary interest in FACES 1997 were Head Start’s impact on children’s development and readiness for school; the quality of education, nutrition, and health services provided to children; the relationships between quality in the classroom and child outcomes; and the program’s impact in strengthening families. One problem with this study, however, is that its design does not allow for comparisons between Head Start participants and demographically similar children who did not attend Head Start. Therefore, although it is possible to conclude that middle-income children continue to outscore Head Start participants, the scope for finer grained conclusions about the gains made by Head Start children in comparison to eligible nonparticipants is quite limited (Barnett & Hustedt, 2003). Results have recently become available for an additional FACES cohort of children who entered Head Start beginning in 2000 (Zill et al., 2003). However, like the initial FACES study, FACES 2000 lacks a comparison group of non–Head Start children. The modest initial effects estimated by Barnett and Camilli (2002) appear to be consistent with results from the FACES studies.

A second large-scale study is the Early Head Start Research and Evaluation project (Love et al., 2002), which began shortly after Early Head Start was established. At the outset of this study, families were randomly assigned to a participant group that received services from the Early Head Start program or to a control group that did not receive these services. Results suggest that this program has a wide range of important short-term impacts, both for 2- and 3-year-old Early Head Start participants and for their parents. When compared to nonparticipants, participating children were less aggressive and more successful on measures of cognitive and language development. Parents of Early Head Start children became more self-sufficient, as they were more likely to participate in job training and educational programs. Furthermore, they showed improvements on assessments of parenting.

Finally, data collection began in 2002 for a third large-scale longitudinal study, which was mandated by Congress during Head Start’s most recent reauthorization in 1998. The National Head Start Impact Study (Puma et al., 2001) will focus on the effects of Head Start participation on children, especially their school readiness, and will also examine the impacts associated with variations in types of services and settings. Children will be followed in this study from age 3 or 4 until the spring of their first grade year. Unlike the FACES studies, the Impact Study employs random assignment of at-risk children to experimental (Head Start) and control (non–Head Start) groups. Although results are not yet available, the use of random assignment gives this study greater promise of producing valid estimates of the effects of Head Start programs on children and their parents. The Impact Study is the most promising evaluation of Head Start’s benefits to date.

**CONCLUSIONS AND IMPLICATIONS FOR FUTURE HEAD START RESEARCH**

Despite the fact that Head Start has been extensively studied since it began nearly 4 decades ago, the answers to some critical questions remain incomplete. The long-term benefits of Head Start rarely have been studied, and never with sufficiently strong research designs. Research on Head Start and similar programs has found substantial long-term benefits in educational achievement and attainment, employment, and social behavior. It is reasonable to conclude that Head Start has positive benefits for school readiness and at least some educational benefits are sustained over time. Less is known about the magnitude of the benefits, the full range of benefits (social, emotional, and physical as well as cognitive), the benefits to parents and siblings of children in Head Start, and the effectiveness of Head Start’s various components.

The strongest evidence for a broad range of large long-term benefits comes from studies of other preschool programs in which key aspects of both the research designs and
programs are of higher quality. This complicates the interpretation of differences in findings between studies of Head Start and other programs. One constant is that initial gains in IQ fade over time. Gains on subject-matter-specific achievement tests are more likely to be maintained. Decreased grade retention and special education rates and increased high school graduation rates are common. Flawed research methods frequently produced results that mimic fade-out with achievement tests, producing an unnecessarily confusing pattern of results. Head Start also seems likely to improve social behavior (e.g., reducing crime), but direct evidence is quite limited.

Across all domains, Head Start’s benefits for children seem likely to be modest in size, smaller than the effects of such well-known interventions as the Perry Preschool and Abecedarian programs. The most obvious reason for the relatively small size of Head Start and Early Head Start effects is the quality and intensity of key aspects of the programs. The programs producing larger effects had much better educated teachers, smaller classes, stronger supervision, and other resource advantages. Head Start’s broad mission to the family may result in smaller impacts on children because its budget is not sufficient to provide intensive services across the board. Head Start Program Information Report data (U.S. Department of Health and Human Services, 2003b) suggest another possible reason for smaller effects. A substantial number of children appear to pass through the program in a given year so that the total number served during the course of the year is considerably larger than the number served for an entire year. And, only about half the children served by Head Start attend the program prior to age 4 enabling them to receive more than a year of Head Start.

An important related question for future research is the relative costs and benefits of the broad array of services mandated to be part of Head Start. Evidence regarding these is spotty, although this includes such bright spots as very large increases in access to dental care (Barnett & Brown, 2000). Results from the Early Head Start study (Love et al., 2002) are promising as they show positive, though small, impacts on a range of outcomes for both children and their parents. In addition, benefits transmitted through parents seem likely to diffuse to siblings, as well. How can the potential to enhance program effectiveness through services to parents be better realized? How can Head Start be reshaped in terms of social, health, and related services? Should its overall budget be increased so that it can increase the intensity of all its services? Should the intensity of selected components be increased, and should this be accomplished by reducing the scope of Head Start’s services and goals?

The past offers some lessons for future research on these questions. IQ tests and their proxies, which include the Peabody Picture Vocabulary Test, may provide reasonable guides to the magnitude of initial cognitive gains, but subject-matter-specific tests are required to more fully assess Head Start’s long-term effects on cognitive abilities. Social and emotional development should be assessed, and studies should not neglect attitudes and behavior in and out of school including motivation, prosocial activities, aggression, delinquency, and crime. Physical development and nutrition seem all too often neglected at a time of increasing concern regarding obesity in children. If such outcomes are neglected in research and evaluation, Head Start policy will be made without much relevant information.

It is particularly important that researchers seek answers to these questions to inform policy decisions that will shape the future of the Head Start program. For example, in 2003 and 2004 Congress considered proposals for a variety of changes in Head Start. The president and others have proposed that Head Start should focus more intensely on literacy, and there are indications that Head Start needs improvement in this area (e.g., McGill-Franzen, Lanford, & Adams, 2002; Zill et al., 2001). However, such a focus could lead Head Start away from an integrated curriculum with broader educational goals and draw down
resources required to provide other services to children and families. The intensive focus on literacy could come at the expense of attention to social and emotional development with dire consequences, given the tremendous potential value of benefits in these domains to Head Start participants and society as a whole (Barnett, 1993; Reynolds et al., 2002). Proposals to be considered also include increasing Head Start teacher education requirements and shifting authority for Head Start to the states. This last raises questions that broader policy studies might answer, such as how funding would be affected by a dilution of federal responsibility for quality and the number of children served.

Although research indicates that Head Start works in its current form, there remains room for research to contribute to further refinement of this program. A research agenda could be developed to study the costs and benefits of Head Start’s components. Another could investigate the effects of enhancements to Head Start 1 component at a time or overall. It would seem especially useful to employ rigorous designs to study improvements to Head Start’s classroom component that would make it more like programs that have produced larger impacts, including increasing teacher quality and compensation, class size, and hours. It cannot be emphasized too strongly that Head Start is a dynamic program that has already been reshaped over the years. Thus, Head Start research provides an historical look at a program that has already adapted to the changing times. It is critical that the research community continue seeking answers to important questions about how to help make Head Start even more successful in the future.

REFERENCES


Lists of current articles: http://depts.washington.edu/isei/iyc/iyc_comments.html

Back to the list of previous articles:

http://depts.washington.edu/isei/iyc/iyc_previous.html