

Evaluating Child and Family Demonstration Initiatives: Lessons from the Comprehensive Child Development Program

Walter S. Gilliam

Yale University Child Study Center

Carol H. Ripple and Edward F. Zigler

Yale University Department of Psychology and Child Study Center

Valerie Leiter

Brandeis University

The Comprehensive Child Development Program (CCDP) was a demonstration project designed to test a specific model of service delivery for young children and families in poverty. Following the evaluation's failure to show strong impacts, early intervention has come under fire from opponents in the popular literature and in Congress. We conclude that shortcomings in both implementation and evaluation contributed to the failure to demonstrate effectiveness. Lessons learned from the CCDP are articulated, addressing the roles of demonstration projects and their evaluations, the problems associated with evaluating programs early in their implementation, the importance of assuring appropriate treatment quantity and quality, and the judicious use and interpretation of large-scale randomized evaluations.

In the summer of 1987, Senator Edward Kennedy introduced legislation to amend the Head Start Act by establishing several centers intended to provide intensive and comprehensive services to children and families living in poverty. The resulting Comprehensive Child Development Act (PL 100-297), signed into law in 1989, authorized for five years the Comprehensive Child Development Program (CCDP), a national multi-site demonstration project designed to test an intervention model that stressed coordinating services for children birth to five

Direct all correspondence to: Walter S. Gilliam, Ph.D., Yale University Child Study Center, 230 South Frontage Road, New Haven, CT 06520-7900; Phone: (203) 785-3384; Fax: (203) 785-7926; E-mail: walter.gilliam@yale.edu.

years old and their families. The goals were to enhance child development and to improve both family functioning and economic status (CSR Incorporated, 1991, December). The program evaluation was conducted by Abt Associates (outcome evaluation) and CSR Incorporated (process evaluation), based on data from 21 of the 34 sites in which CCDP was implemented.

Against the background of Congressional support and an apparently sound theoretical base, the impact evaluation findings by Abt Associates (1997, June, 1997, September) released by the Administration for Children and Families in January 1998 were sobering. Abt concluded that after five years of operation CCDP had virtually failed to produce significant gains among participants.

The evaluation results have been referenced in policy-oriented discourse about the effectiveness of other federal government programs for poor children and families in general. For example, the CCDP results were presented by Abt at a meeting at the American Enterprise Institute for Public Policy Research where the future of Head Start was discussed. Suggesting that the findings demonstrated the ineffectual nature of federal programs, the results were used to bolster the case for placing Head Start and other federal programs under the control of the states. (For more information about the current debate regarding the devolution of Head Start, see Ripple, Gilliam, Chanana, & Zigler, 1999.) In addition, the CCDP findings have been used by Douglas Besharov, a well-known policy analyst, to question the value of Early Head Start. The Abt findings were entered as testimony in a joint Senate/House subcommittee hearing on the topic of Head Start's effectiveness (St. Pierre, 1998, March 26). Referring to Early Head Start, the brief that was circulated before the hearing stated: "The closest research on the effectiveness of similar programs is the Comprehensive Child Development Program. A recently released study by Abt Associates found the program to have had virtually no impact on program participants or their families" (Head Start: Is It Making a Difference? Can It Be Improved?, 1998, March 26).

The popular media echoed this broad interpretation of the CCDP evaluation findings soon after their release. In an article appearing in *Newsweek*, the CCDP findings were used to argue that *all* publicly supported programs designed to help children and families in poverty were a waste of time and money (Samuelson, 1998, February 23). Because the CCDP evaluation had failed to detect measurable gains, it was argued that no federal programs for poor children and families could provide meaningful benefits. The article concluded that "large federal programs, whatever their benefits, can't undo parental failure. Nor can they offset the ill effects of family breakdown" (p. 45).

In the context of over 30 years of research on the child and family effects of early childhood intervention programs (Barnett, 1995; Bryant & Maxwell, 1997; Guralnick, 1997; Haskins, 1989; Meisels & Shonkoff, 1990), this over-generalization of the CCDP evaluation results is ill founded. How, then, should the Abt results be interpreted? Particularly because the findings have been used to challenge other efforts with similar goals but dissimilar models, such as Early Head Start, thoughtful scrutiny of the findings and their implications is critical.

The purpose of this paper is to glean constructive lessons from the CCDP impact evaluation in a way that moves forward the fields of early intervention and

program evaluation. We explore several questions regarding CCDP and its evaluation: What do the evaluation findings tell us about the CCDP model, specifically about the role of family support in early intervention programs? What can be learned from the one most successful program site? Which aspects of the evaluation itself contributed to the null findings? How can subsequent evaluations of demonstration programs be improved? Before addressing these complex issues, we briefly describe the CCDP model, the evaluation methodology, and the results.

WHAT WAS CCDP?

First and foremost, CCDP was a national demonstration project. Demonstration projects—ideally used to explore the usefulness of programs prior to their full implementation—are experiments designed to test ideas. They have the potential of identifying programs that can be of enormous benefit to society, as well as those that might represent a relative waste of energy and resources, if broadly implemented. Programs that are shown to be successful may justify expansion; those that are not successful warrant discontinuation or significant revision. CCDP's purpose as a demonstration project was to determine the impact and cost-effectiveness of the CCDP service delivery model, as described later in this paper.

Contextual Background

The CCDP model was inspired by several existing early childhood programs, including Head Start, Parent Child Centers, the Child and Family Resource Program, and the Center for Successful Child Development (the Beethoven Project) in Chicago. Building on these programs' theoretical bases and perceived strengths and weaknesses, CCDP was designed to create a comprehensive system of intervention relying heavily on case management and consisting primarily of referral to extant programs in the community (e.g., social service agencies, day care programs, etc.), rather than providing services directly.

The optimism behind CCDP was not without basis. Well-conceptualized primarily parent-focused home visiting programs can yield positive impacts for parents, whereas impacts on children are less robust (Olds & Kitzman, 1993; Seitz & Apfel, 1999). Although the target population, scope of services, or desired impacts may differ from those of CCDP, several parent-focused interventions have documented at least some impacts on either parents or children. For example, the Parents As Teachers Program (PAT), a home-visiting curriculum designed to help parents provide a strong foundation of school readiness for their young children (Winter & McDonald, 1997), has shown some positive results for children in two randomized demonstrations. Findings suggested that PAT had some effect on parent-rated child development, although results were mixed and differed as a function of ethnicity (Wagner, 1998). Another parent-focused, home-based program, the Yale Child Welfare Program (Provence & Naylor, 1983), targeted poor women expecting their first child. The treatment lasted until the children were 30 months old and consisted of medical, developmental, social,

and parent vocational services. Results indicated that participant children, as compared to matched time-lag controls, exhibited significantly greater language development (Seitz & Provence, 1990). Several positive outcomes were found to persist in a 10-year follow-up of program participants (Seitz, Rosenbaum, & Apfel, 1985).

Possibly the clearest evidence of an effective parent-focused home-based program is from the Elmira Nurse Home Visitation Study (Olds et al., 1997). Four-hundred economically disadvantaged women, pregnant with their first child, were randomly assigned to a no-treatment control group or a treatment group that received two years of extensive home visitation services delivered by well-trained medical staff. The results revealed that treatment was associated with significant reductions in welfare dependence, maternal criminality, child abuse and neglect, and rapid subsequent pregnancy (Olds, Henderson, Chamberlain, & Tatelbaum, 1986). Results of a larger replication study support these initial findings (Kitzman et al., 1997).

The CCDP Service Delivery Model

CCDP legislation defined it as a comprehensive program for low-income children birth to five years old and their families (Abt Associates, 1997, June, 1997, September). The legislative goals were to enhance child development both indirectly (by providing parenting education and economic development to the parents) and directly (by providing quality early childhood experiences and preventive health care to the child).

Despite its comprehensive intent, CCDP was implemented primarily as a case management program, consisting mostly of a family needs assessment, 30- to 90-minute weekly home visits, arrangements for housing and utilities assistance, and optional referrals to a variety of community services (Abt Associates, 1997, June). For example, caregivers might be referred to educational and employment services (e.g., adult education, vocational training, job development and placement, life skills education, and collateral child care), health and prenatal care, and substance abuse services. The majority of these services were directed toward parents and were primarily vocational, with the intent of helping them secure employment. For younger children aged birth to 3 years old, services included optional home-based parenting education (for approximately one hour every two weeks). For children 3 to 5 years old, optional center-based child-care programs were offered. Although child care was offered, comprehensive early childhood education programs, such as Head Start, were not a typical component (Abt Associates, 1997, June).

After assessing family needs, case managers typically referred caregivers and children to existing community programs for services, and paid for services when no other mechanism of payment existed. When there were no extant programs in the community, CCDP provided the needed services directly. Thus, CCDP relied heavily on two factors: (a) the availability and quality of services previously extant in each community; and (b) the skills of the case managers at each CCDP site. Case management services were provided by paraprofessionals, whose level

of training and experience in providing two-generation focused case management and referral services has not been reported. In essence, CCDP was not implemented as a unique, clearly articulated early intervention program. Rather, it was a service brokerage system designed to identify specific family needs and then refer out for services, particularly as these identified service needs related to the government mandate for economic self-sufficiency (Abt Associates, 1997, June).

Evaluation Methodology

Families from the 21 CCDP evaluation sites were below the current federal poverty level, had a focus child under one year old or unborn, and agreed to participate in CCDP for five years. At the time of recruitment, modal annual household income for participants was \$4,000 to \$4,999, and the modal age of the focus child was 1 to 2 months, though services were actually begun somewhat later. Families represented a variety of ethnic and regional groups, and were randomly assigned to treatment ($n = 2,213$) and control ($n = 2,197$) groups. Evaluation assessments were conducted near baseline and at regular intervals throughout the five-year intervention. Data included measures of child developmental status, parenting beliefs and values, and family self-sufficiency.

Results of the Evaluation

Cross-sectional and longitudinal analyses were conducted to test CCDP's impact on children's cognitive development, social-emotional development, and physical health. Caregiver outcomes included maternal self-sufficiency, life management skills, and psychological and physical health (Abt Associates, 1994, May). Overall, only a few positive impacts were noted at the end of the five-year program, as described below. In all cases, effect sizes were minimal.

Effects on parents. The emphasis on vocational services for the parents paid off: most significant impacts involved employment and economic benefits for the mothers. At the end of the program, CCDP mothers were significantly more likely than controls to have been enrolled in some form of academic, vocational, or job training ($p < .01$; 26% versus 22%, respectively). Also, CCDP had a slight, but statistically significant, impact on the rate of increase in household income over the duration of the project. CCDP mothers also reported less belief in the need for corporal punishment than did control mothers ($p < .05$; Abt Associates, 1997, June).

Effects on children. CCDP children showed a slight but statistically significant advantage over controls at age two years, in terms of their overall developmental level on the Bayley Scales of Infant Development (Abt Associates, 1994, May); however, these differences were not found in subsequent years. Overall, as is typical among poverty samples, developmental and cognitive scores were in the below average range for both groups. CCDP parents rated their children slightly better developed at the end of the program ($p < .01$). No significant differences were noted in preventive health care, child mortality, and

birth characteristics of subsequent siblings. As a caveat, parent report measures were often revised or reorganized in ways that may have compromised their validity. For instance, one parent-rating test was used for children older than those for whom it was intended, and two tests were revised in some unspecified way.

An Effective CCDP Site

Although the overall results of the CCDP evaluation were disheartening, one of the 21 evaluation sites did produce consistent, statistically significant impacts. Families at this site experienced positive effects in several key domains, typically of moderate effect size (Abt Associates, June 1997, pp. 6-7-6-8). Specifically, during their participation in the program, CCDP children's scores on the Peabody Picture Vocabulary Test improved by over nine standard score points, and their scores on the Kaufman Achievement Battery for Children improved by nearly four points. In terms of self-sufficiency, mothers or partners in the family were 22% more likely to have been employed; 20% fewer mothers received AFDC at the end of the study; and on average, CCDP families received Food Stamps for 19% less time. The CCDP families also experienced a significant increase in annual family income (\$3,622 more per year). Finally, CCDP parents at this site became more aware of their children's needs and had more appropriate expectations of their children. Based on the Adult-Adolescent Parenting Inventory scales, over half (54%) of the control group parents were at risk of abusive behavior toward their children, compared to 33% of the CCDP parents. Without appropriate inter-site data, we may only hypothesize why this site was effective. However, it appears that several factors may have accounted substantially for the effectiveness of this CCDP site.

First, the program was well organized and implemented from the time it served its first cohort of families. The staff at this site had a great deal of previous social-work experience with child-development programs in the host community, and the Project Director had 17 years of experience at this site (Project Director, personal communication, May 6, 1998). Starting in 1985, they operated a home-based program for infants and toddlers, providing a solid model for CCDP implementation. Whereas most sites experienced high rates of administrative turnover and reorganization (CSR Incorporated, 1997, March), this site's leadership was stable for the entire CCDP program. They routinely collected data on operations, and used them to improve their program.

Second, this CCDP site was thoroughly integrated into the community long before formal CCDP operations began. The Project Director and other management staff had served on community boards outside of the program, and the prior home-based program allowed staff to put a community advisory committee into place before even applying for the CCDP grant (Project Director, personal communication, May 6, 1998). After CCDP operations began, this site expanded community involvement in the program by creating additional advisory committees. Of the 21 CCDP study sites, this one was the only one located in the public schools (Abt Associates, 1997, June). This may have strengthened the institutional link between CCDP and the school, and there is some evidence that comprehen-

sive programs implemented as part of a public school program can be highly effective (Dryfoos, 1994; Zigler, Finn-Stevenson, & Stern, 1997).

Third, the families at this site differed from those at other sites in a number of ways. The site was located in a small city in a rural area. Relative to CCDP families at other sites, families here were more likely to have a partner in the home (59% vs. 38% overall); the focus children were less likely to have teenage mothers (26% vs. 36% overall); and the families had higher *per person* annual incomes at baseline (\$2,390 vs. \$1,780 overall). Thus, families at this site were somewhat less at risk, which may account at least partially for this site's greater effectiveness.

WHAT WENT WRONG?

Despite the success of the one site described above, the overall evaluation results for CCDP were disappointing. Several aspects of the model and the evaluation methodology may have contributed significantly to these overall lackluster findings. Whereas issues we raise may be self-evident to program evaluators, examining them in this particular context is essential given their political and social policy importance.

The Evaluation Was Too Early

The CCDP results were based on data from the first cohort of children served by this fledgling program. Thus, the evaluation could not help but reflect a significant amount of "start-up noise." It takes some time to create an infrastructure for service delivery; recruit, train, manage and maintain staff; establish essential community links; promote awareness of a new program; and work out various operational problems that become apparent only after a program opens its doors. As the sites matured and stabilized, the number and intensity of the services new CCDP families received increased (CSR Incorporated, 1997, March). For example, staff at one site noted that it took at least two years after start-up to significantly reduce dropout and provide the full complement of quality developmental, physical health, and mental health services to children and vocational services to parents. These programmatic improvements translated into better outcomes for the families who enrolled after this two-year implementation phase (Striefel & Robinson, 1998, February).

Because the evaluation used data from only the first cohort of CCDP families, the effect of program maturation and its desirable results for subsequent cohorts were missed in the evaluation. In other words, poor initial implementation of CCDP may have led to poor initial service utilization among participants, leading to increased early program dropout.

The Random Assignment Appeared to be Less than Random

Random assignment to treatment and control groups was not centrally controlled and resulted in several problems that raise questions about the compara-

bility of the groups. At some sites, random assignment was conducted by the programs themselves, whereas at other sites it was overseen by CSR, the group responsible for the process evaluation (Abt Associates, June 1997). In spite of attempts at randomization, many pre-treatment differences were noted between these groups on several key baseline demographic variables (ethnicity, single-parent homes, maternal employment, maternal high-school degree, teen parents, low birth weight, and income). Unfortunately, these differences were only reported on data at the site level and not for the overall sample. Eleven significant differences were reported on these seven different baseline variables at 21 sites ($p < .05$; Abt Associates, June 1997), somewhat higher than the seven significant differences one might expect from chance alone ($7 \times 21 \times .05 = 7.35$). In terms of percent of maternal employment, treatment and control groups differed significantly ($p < .05$) at 4 of the 21 sites on this variable alone (a rate nearly four times greater than expected by chance alone). Unfortunately, it is not clear how to interpret these baseline differences, since neither the direction nor magnitude of these differences were reported. Nonetheless, in this case, it appears that leaving subject assignment to the service providers at the individual sites may have compromised the randomization, and thus the comparability of the groups.

Controls Received Services Similar to CCDP Participants

An essential assumption of program evaluation is that the experimental group receives the treatment and the control group does not. This assumption was largely unmet in the CCDP evaluation. Given the high dropout rates in the start-up cohort, the case management strategy was largely unsuccessful in fostering better service utilization among first cohort CCDP families. In addition to generally low treatment levels among CCDP families, children and families in the control group obtained similar levels of treatment in the community (CSR Incorporated, 1997, March; Layzer, St. Pierre, & Wilwerding, 1998, July). The primary difference between the two groups was that CCDP families were more likely to receive some form of ongoing case management. The only other differences were that CCDP mothers were about 40% more likely than controls to receive some amount of adult education and vocational/job training, and CCDP focus children received about one hour more of center-based child care per week (Abt Associates, 1997, June; CSR Incorporated, 1997, March). If these were the only differences in service utilization between the two groups, it would be very difficult to argue that the CCDP families received significantly more services than did controls. Also, as many as 75% of recruited families in both groups were already identified by social service programs, and may have received services prior to CCDP intake (Abt Associates, 1997, June). This over-reliance on recruitment by social service referrals may have contributed to the similarities in service utilization between CCDP and control families. Nonetheless, the evaluation failed to consider these treatment similarities when assessing the impact of CCDP.

Program Dropout Was High, and Participation Rates Were Uneven

Very few CCDP families in the first cohort received all five years of CCDP, but all were included in the final impact analyses. About 5% of CCDP families dropped out before receiving *any* CCDP services, yet they, too, were included in the evaluation as having participated fully in CCDP. Furthermore, about 13% of the original sample dropped out during each year of the program. Only one-third of the CCDP families were enrolled the full five years, and less than one-half completed four years (Abt Associates, 1997, June; CSR Incorporated, 1997, March). These figures are even more disheartening when one considers that the recorded date of service termination was typically up to six months after families actually stopped receiving services. Also, service delivery began some time after families were officially enrolled in the program. In other words, the above figures may overestimate the length of services that CCDP families received by over one-half year (Abt Associates, 1997, June).

Beyond program duration, the level of families' actual participation was often low: many families did not consistently or actively participate in services. Hoping they would return to a more active level of participation, CCDP staff were often unwilling to discontinue services to these families (Abt Associates, 1997, June). As a result, CCDP families received drastically dissimilar services, in terms of length, intensity, and quality. In most cases it was difficult to determine exactly what the CCDP families received. When it was possible to determine what they received, it was typically far less than what would be expected from the model. Despite high attrition rates and many unanswered questions regarding CCDP participation and offerings, all families were retained in the study as if they had received the full program. These problems were further complicated by poor response rates on evaluation measures (60% to 80%, depending on the type and time of data collection; Abt Associates, 1997, June). Abt did attempt to address these issues in the analyses, but these attrition rates were too high to be statistically treated. Even reanalyzing the data for participants who completed the full course and participated in various services would have required precise data on service access and participation rates. Instead, it would have been far better to evaluate CCDP impacts after program attrition rates stabilized.

With these rates of attrition and poor participation, it is impossible to estimate the effects the program might have had with reasonably high levels of participation. CCDP's high reliance on social service referral as a means of recruiting subjects might conceivably have contributed to poor participation. If recruitment had been driven by active participant self-referral, rather than passive referral by other services, participant investment might have been higher and positive impacts more possible.

High-Quality, Child-Focused Services Were Under-Emphasized

The CCDP model, which was primarily parent-focused, may have under-emphasized the importance of high-quality services administered directly to children. Although data support the importance of involving parents when inter-

vening in the lives of children (Smith, 1995), mounting empirical evidence suggests that the most effective way to impact children is by providing high-quality, child-focused services (Barnett, 1995; Yoshikawa, 1995). Depending on the age of the children, only 19% to 22% of CCDP participants attended early childhood education programs, such as Head Start, as did 11% to 22% of controls (Abt Associates, 1997, June). Furthermore, average attendance was only about two days per week for preschoolers and about one day per week for younger children (CSR Incorporated, 1997, March). Overall, children in both groups attended a variety of child-care programs of unknown quality, and neither group appeared to have consistently received high quality early childhood education.

Mental Health Services Were Under-Emphasized

Unaddressed mental health problems may contribute to parents' difficulties in participating in work, education, and treatment programs (Gelfand & Teti, 1990), and children of depressed mothers are at increased risk for developmental and social-emotional problems (Beardslee, Bemporad, Keller, & Klerman, 1983; Downey & Coyne, 1990; Sameroff, Seifer, Barocas, Zax, & Greenspan, 1987). Following the release of the final evaluation report, Abt investigators performed analyses in which maternal depression was shown to have a negative relationship to parenting attitudes, risk for child abuse and neglect, quality of parent-child interaction, employment, and economic self-sufficiency (Layzer et al., 1998, July). However, mental health services appear to have been under-emphasized by CCDP. Although 42% of CCDP mothers screened positive for depression, only 15% received mental health services (Abt Associates, 1994). Furthermore, the average number of mental health contacts for those who did receive mental health services was only 4 to 6 per year (CSR Incorporated, 1997, March), and no data are reported regarding the modality or quality of the services that were received. Service utilization rates were not reported for children or for the control mothers, 42% of whom also screened positive for depression. Maternal depression was acknowledged in the interim report (Abt Associates, 1994), but then was not included in the final analyses of the program's impact (Abt Associates, 1997, June), except as it was addressed well after the final evaluation report was made public (Layzer et al., 1998, July).

LESSONS LEARNED FROM THE CCDP DEMONSTRATION PROJECT

What can be learned from what went wrong? Beyond the disappointing findings, there is much to be learned from the CCDP demonstration and evaluation regarding early intervention program design and evaluation in general. Many of these lessons revisit basic principles of program evaluation.

Lesson 1: Demonstrations are Experiments First and Foremost

The goal of demonstration projects is to determine which programs work and should be implemented, versus which programs do not and should be either revised or phased out. For example, Home Start, implemented in March of 1972, was a success and continues today. The Home Start model, essentially a home-based Head Start program, was demonstrated at 16 program sites, serving 2,500 children. After 3½ years, an evaluation revealed that Home Start was as effective as Head Start, based on impacts on children's school readiness, social-emotional development and medical and dental care as compared to non-treatment controls (Love, Nauta, Coelen, Hewett, & Ruopp, 1976, March). Based on this evidence, Home Start was expanded and as of FY 1996 was being implemented at 571 sites nationwide and serving 39,833 children (about 5% of the total Head Start population) (Administration for Children and Families, 1998). In contrast, the 29-site national demonstration project Health Start was not a success. Health Start's goal was to provide medical services to poor children during a brief summer program, but the program was too brief to accomplish the desired goal and failed (Valentine, 1997). It is expected that some demonstrations will succeed and others will fail, as the ultimate objective of any demonstration project is to learn what works best.

Lesson 2: External Validity Matters

External validity refers to the degree to which one can assume that the findings of a particular study can be generalized to the population from which the sample was drawn. This depends on the degree to which the study sample represents the population to which the findings might be generalized. In a national multi-site demonstration project this depends on the degree to which the program sites represent the nation. CCDP was evaluated in sites that represent many different geographic regions of the nation, and the study sites were both primarily rural and urban (Abt Associates, 1997, June). Despite this, it is not clear that these sites truly represent the United States, as they were selected based on their community's ability to deliver a variety of services. Therefore, it is not certain that the evaluative findings from these sites are applicable to communities where fewer services and case management resources are available.

External validity in program evaluation might also refer to the degree to which evaluative findings for one program might generalize to other similar programs. In order to place evaluative findings in context, social scientists may use the results of program evaluations to make inferences about other programs that they see as similar (Barnett, 1995; St. Pierre & Layzer, 1998). Extrapolating one set of findings to programs that are very different in nature, scope, or intensity, however, is ill advised. Although it may be tempting to generalize the results of the CCDP evaluation to two-generation programs that provide direct service to children and their families (St. Pierre & Layzer, 1998), the findings may be applied more accurately to other case-management systems of care. One may be able to argue for the applicability of the CCDP findings to other primarily adult-focused

case-management programs, but clearly not for child-focused, or even two-generation-focused programs that provide direct services.

Lesson 3: Evaluations Should First Measure the Intended Outcomes

Identifying the intended outcomes of a complex program is not always straightforward. Because the majority of services were not provided directly, the quality, intensity, and resulting impact of these services were outside of the control of CCDP. As a service brokerage system, CCDP should have been evaluated first on its ability to achieve what it was implemented to do (i.e., link families to needed services). Indeed, CSR did evaluate this and found that CCDP did not significantly increase service utilization rates in its first cohort (CSR Incorporated, 1997, March). This being the case, it makes little sense to evaluate the impacts of services that were largely not provided. Once it was learned that treatment families were not receiving more services than controls, the impact evaluation should not have even been attempted.

In conceptualizing intervention programs, program goals should dictate program design and implementation. Program implementation, in turn, should produce the desired outcomes. CCDP's *goals* were to directly impact child development and parent/family self-sufficiency, suggesting the need for an impact evaluation based on direct outcomes on child development and family functioning. However, CCDP as *implemented* was a service utilization program that was heavily weighted toward adult educational and vocational services. In essence, CCDP's goals and evaluation (direct impact on child development and family functioning) did not match its implementation design (a case referral system for primarily welfare-to-work oriented community services).

Lesson 4: Implementation Takes Time

Often, when an innovative demonstration program attracts political interest, the pressure to fully implement the program is immediate. Head Start is a case in point: although psychologists suggested that Head Start begin as a pilot program serving a manageable 2,500 children and then expand after an evaluation of its effects, political pressure exploded it to a full-fledged program for more than half a million children in its first summer (Zigler & Anderson, 1997).

New initiatives are also pressured to produce data showing strong benefits early in the life of the program. As noted, this was the case with CCDP, and relying on the first cohort's data all but assured that the evaluation measured start-up noise as much as anything else. In addition, evaluating too soon may not allow sufficient time to pilot the outcome measures, and to work out problems with the evaluation itself.

Too often, when outcome evaluation commences with the start-up phase of a program, process evaluation is forgotten altogether. Sound program evaluations follow a pilot or implementation phase with a process or theory-driven evaluation, and only then the collection of outcome data. In this sequence, process data can inform the development of both the program itself and of the evaluation focus and

methodology. The CCDP evaluation did include a process report, but the outcome data were collected concurrently, and both began near the outset of program implementation. By pursuing process and impact evaluations concurrently, the potential benefits of a process evaluation were lost. CCDP was therefore essentially unable to utilize the process findings to improve services and ultimately measured impacts, unlike the one successful site previously mentioned.

Rather than expecting immediate overwhelming results, innovative programs should be approached with the goal of developing a model based on existing knowledge, implementing that model, studying it, and then modifying the model accordingly (Campbell, 1969). Once the model has been well implemented, as established by a process evaluation, then outcome data can be collected and used to modify the program further, to achieve maximum effectiveness. The results of the one effective CCDP site suggest that the CCDP model *can* work. Unfortunately, there was no time or mechanism to allow the success of that one effective site to inform and improve the other sites.

Of course, the push for program accountability is intended to save money by weeding out ineffectual programs. Yet, the resulting emphasis on early evaluation can backfire as agencies feel pressured to evaluate new programs too soon. So long as policy-makers demand immediate—rather than timely—evidence of accountability, and so long as scientists accommodate these unrealistic demands, programs will continue to be evaluated before they are ready. The net result is wasted evaluation money, the risk of inaccurate results, and the premature end to efforts that might have been of tremendous benefit had they been allowed time to mature.

Lesson 5: Treatment Compliance Cannot Be Assumed

The CCDP population of low-income, often teenage mothers, is a difficult one to serve. Of course, there is no way to force volunteers to avail themselves of services, just as there is no way to ensure that control subjects will not obtain other comparable treatments. However, the failure of the CCDP evaluation design to include variables to fully account for treatment duration and intensity renders the findings moot. Clearly, *enrolling* and *participating* are not the same thing. Data on who does participate, and, among those, who improves, would be valuable. This point is particularly salient in light of the data suggesting relatively high rates of maternal depression among CCDP participants. The consideration of important moderating variables, such as maternal depression or substance abuse, is key to assessing treatment compliance and judging program effectiveness.

Lesson 6: Quality Counts

Service quantity and quality are linked to outcomes, but without some form of quality assessment the provision of high quality services cannot be assumed (Berlin, O'Neal, & Brooks-Gunn, 1998; Frede, 1995). For example, although the CCDP model provided some type of early childhood program, the quality of the services available in the communities was apparently not taken into account. Even

though CCDP children attended child care for slightly more time than comparison children, better child outcomes could not be anticipated without knowing something about the type and quality of the programs they attended. The same can be said about the other program components as well: quality is particularly important to assure—and to document—when service delivery is based on a model of pre-existing services. The CCDP model is remarkably similar to the Fort Bragg Demonstration Project model (Bickman et al., 1995). Fort Bragg, too, failed to show positive outcomes, and although the failure likely stemmed from a variety of factors (Pires, 1997) the assurance of quality treatment was identified as a major issue (Bickman, 1997).

Despite Abt Associates' assurance in its final report that "CCDP projects were well implemented by local grantees" (Abt Associates, 1997, June, p. 8-9), there is no compelling evidence that program quality was measured in a meaningful way. Abt rationalized their assessment of program quality by stating that "CCDP served the families that it was intended to serve, coordinated the efforts of thousands of service agencies nationwide, and delivered a wide range of services to a high proportion of participating families" (Abt Associates, 1997, June, p. 8-9). However, as we have shown, the evidence does not equate these three assertions with quality. First, although CCDP did serve its intended population, "serving" and "serving well" are two very different issues. Second, coordinating services does not mean that the services were high quality. Third, the "high proportion of participating families" was only about one-third, and that did not take into account late starts, early retreats, and poor treatment compliance. In sum, there is virtually no evidence to suggest that CCDP services were of sufficient quantity or quality to effect positive changes.

Lesson 7: Large Randomized Studies Should Not Be the First Step in Evaluating New Programs

Random assignment of participants to treatment and comparison conditions is widely acknowledged as the best methodology for research studies in general and evaluation studies in particular. Its purpose is to help ensure that the experimental and control groups do not significantly differ on important quantifiable and non-quantifiable variables. The difficulties involved with this methodology, however, often become apparent in "real-world" effectiveness studies. Assignment to a treatment or control group does not ensure that subjects will comply with their placement (Heckman & Smith, 1993, May), just as asking service providers to randomly assign CCDP evaluation families did not ensure equivalent groups. Furthermore, in many cases, random assignment to a no-treatment group is complicated by ethical issues (e.g., the withholding of treatment) and the difficulty involved in finding true "no-treatment" control groups (Advisory Panel for the Head Start Evaluation Design Project, 1990).

Randomized evaluations of complex programs such as CCDP are similar to open clinical drug trials in medicine (Prien & Robinson, 1994). In open trials, large groups of people often are randomly assigned to either experimental or control groups, and are prescribed a specific drug that is being studied in order to

estimate what its effectiveness will be after it has been marketed. Strict treatment compliance by participants in the experimental group is not demanded, and controls are free to seek other treatments. Such being the case, the control group is far from a no-treatment group. Rather, it may be more accurately conceptualized as representing a *usual-treatment* group. This was the case with the CCDP evaluation: the control group was a usual-treatment group in that they obtained usual community services. The lack of a no-treatment control group is not so much a flaw, as it is a limitation that indicates that the results can only be interpreted as a study of program effectiveness vis-à-vis extant services (whether the program significantly adds anything to existing services) and not a study of program efficacy (whether the program would work in its intended intensity and in the absence of other services). With CCDP, as with many other social programs, open trials were begun before any preliminary data were obtained regarding the effectiveness of the model when administered in its intended quality and intensity, and as compared to no services. As a result, the reason for the overall null findings cannot be ascribed with any certainty either to CCDP's model and design or to its large-scale implementation, and little can be done to determine and then address CCDP's shortcomings.

CONCLUSION

The early intervention community has been remarkably silent in the wake of the Abt Associates' CCDP outcome evaluation report. The program that was meant to provide a "super Head Start" was reported to have largely failed to deliver the expected positive outcomes, and the frightening implication of that failure was that helping low-income families was a hopeless task unworthy of taxpayers' support. Indeed, the threat implied here for other federal programs—particularly for Early Head Start—was tangible.

How are the findings best understood? Surely, not by erroneously extrapolating the null findings from this one outcome study to all other intervention programs. To suggest that the failure to show positive results from this one demonstration means that child-focused and two-generation programs cannot work is to erroneously over-generalize these findings to programs that are very different in model. Considered another way, the results may suggest that two-generation programs that serve both parents and children directly, such as Early Head Start, may be a *wiser* investment than programs that primarily serve caregivers only.

Perhaps the most valuable lessons to be learned from the CCDP demonstration regard the way large-scale evaluation projects are conceptualized and implemented. Whereas random assignment plays an important role in evaluating the impact of social programs, randomization, in and of itself, does not result in a well-conducted impact evaluation. Complex social programs require a complex array of evaluative techniques, and transactional models that assess the fit between treatment and participant are under-utilized. Equally important, however, is ensuring the integrity of the experimental variable, in terms of its duration, intensity,

and quality. Validating the independent variable is only accomplished by a thorough process evaluation before the collection of outcome data. It is difficult to imagine discovering positive impacts when two-thirds, if not more, of the treatment group failed to receive the treatment in its intended dose, and when the control group often obtained similar services on their own. Yet, this was the case with the start-up cohort of the CCDP evaluation. Why did this happen? Because randomly assigning people to a “treatment” or “no-treatment” group does not ensure that they will comply with their placement. However, if one compares those who actively pursue treatment to those who do not, the desirable condition of random selection is lost and the experimental variable is biased by motivational and other issues. At first, it may appear that randomization and treatment integrity are opposing conditions: when one is ensured, the other is lost. The issue is most problematic during a program’s start-up cohort, when treatment duration, intensity, and quality have not yet been maximized.

One solution to the randomization versus treatment-integrity dilemma, inherent in all social program evaluations, is to pursue process and outcome evaluations sequentially, rather than concurrently. Outcome evaluations should not be attempted until well after quality and participation have been maximized and documented in a process evaluation. Although outcome data can determine the effectiveness of a program, process data determine whether a program exists in the first place. If a process evaluation determines that attrition rates are high and that important services, such as mental health, are under-utilized, the program is not yet ready for an expensive impact evaluation that is likely to yield null findings. Rather, process evaluation findings should be used to help improve the program and move it closer to a point when an impact evaluation is appropriate. The desirable serial sequence of program evaluation is: small-scale efficacy studies, process evaluation of large-scale implementation, impact evaluation, then cost-benefit analysis. Failure at any one level should preclude continuing to more advanced levels of program evaluation. Although policymakers likely will continue to demand immediate impact findings, it remains the responsibility of scientists to educate research consumers about the need for thoughtful, timely assessment, as well as the limitations of rushed evaluations.

Considerable time, money, and expertise went into CCDP. Was the demonstration a failure? Although CCDP has completed its legislative purpose as a demonstration project and has been terminated, the demonstration itself cannot be considered a failure so long as researchers and policymakers glean lessons from the results. Indeed, CCDP became a major consideration in the design of Early Head Start. Because poor families in the United States are facing deepening poverty, increasing risk factors, and struggles with welfare reform, developing and testing new intervention strategies, as well as improved methods of evaluating them, are crucial to ensuring the healthy development and adjustment of future generations of children. Despite the CCDP evaluation’s disappointing results, the lessons learned should at least aid in improving the conceptualization, implementation, and evaluation of future demonstration programs for low-income children and their families.

Acknowledgment: The authors express grateful appreciation to Dr. William Harris. His erudite counsel and extraordinary enthusiasm substantially contributed to the process and product of our efforts. The authors also acknowledge the helpful comments and critiques offered by several colleagues, in particular Dr. John M. Love, and the anonymous reviewers at *Early Childhood Research Quarterly*.

REFERENCES

- Abt Associates. (1994, May). *National impact evaluation of the Comprehensive Child Development Program: Interim report*. Cambridge, MA: Author.
- Abt Associates. (1997, June). *National impact evaluation of the Comprehensive Child Development Program: Final report*. Cambridge, MA: Author.
- Abt Associates. (1997, September). *The effectiveness of comprehensive case management interventions: Findings from the national evaluation of the Comprehensive Child Development Program*. Cambridge, MA: Author.
- Administration for Children and Families. (1998). Head Start 1998 fact sheet. Available at: <http://www.acf.dhhs.gov/programs/hsb/facts98.htm>. Accessed. February 2, 1999.
- Advisory Panel for the Head Start Evaluation Design Project. (1990). *Head Start research and evaluation: A blueprint for the future*. Washington, DC: Administration for Children, Youth and Families.
- Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *The Future of Children*, 5(3), 25–50.
- Beardslee, W. R., Bemporad, J., Keller, M. B., & Klerman, G. L. (1983). Children of parents with major affective disorder: A review. *American Journal of Psychiatry*, 140, 825–832.
- Berlin, L. J., O’Neal, C. R., & Brooks-Gunn, J. (1998). What makes early intervention programs work? The program, its participants, and their interaction. *Zero to Three*, 18(4), 4–15.
- Bickman, L. (1997). Resolving issues raised by the Fort Bragg Evaluation. *American Psychologist*, 52, 562–565.
- Bickman, L., Guthrie, P. R., Foster, E. M., Lambert, E. W., Summerfelt, W. T., Breda, C. S., & Heflinger, C. A. (1995). *Evaluating managed mental health services: The Fort Bragg experiment*. New York: Plenum Press.
- Bryant, D., & Maxwell, K. (1997). The effectiveness of early intervention for disadvantaged children. In M. J. Guralnick (Ed.), *The effectiveness of early intervention* (pp. 23–46). Baltimore, MD: Brookes.
- Campbell, D. T. (1969). Reforms as experiments. *American Psychologist*, 24, 409–429.
- CSR Incorporated. (1991, December). *Comprehensive Child Development Program—A national family support demonstration: First annual report*. Washington, DC: Author.
- CSR Incorporated. (1997, March). *Comprehensive Child Development Program—A national family support demonstration: Process evaluation final report*. Washington, DC: Author.
- Downey, G., & Coyne, J. C. (1990). Children of depressed parents: An integrative review. *Psychological Bulletin*, 108, 50–76.
- Dryfoos, J. G. (1994). *Full-service schools: A revolution in health and social services for children, youth, and families*. San Francisco: Jossey-Bass.
- Frede, E. C. (1995). The role of program quality in producing early childhood program benefits. *The Future of Children*, 5(3), 115–132.
- Gelfand, D. M., & Teti, D. M. (1990). The effects of maternal depression on children. *Clinical Psychology Review*, 10, 329–353.
- Guralnick, M. J. (1997). Second-generation research in the field of early intervention. In M. J. Guralnick (Ed.), *The effectiveness of early intervention* (pp. 3–20). Baltimore, MD: Brookes.
- Haskins, R. (1989). Beyond metaphor: The efficacy of early childhood education. *American Psychologist*, 44, 274–282.
- Head Start: Is It Making a Difference? Can It Be Improved? (1998, March 26). *Tentative Witness*

- List & Issues for the 105th Congress* [Limited Circulation]. Joint Hearing of the United States Senate Subcommittee on Children and Families and the U.S. House Subcommittee on Early Childhood, Youth and Families (Senate Hearing 105-553).
- Heckman, J. J., & Smith, J. (1993, May). *Assessing the case for randomized evaluation of social programs*. Paper presented at the meeting of the Danish Presidency Conference, "Effects and measuring effects of labour market policy initiatives," Kolding, Denmark.
- Kitzman, H., Olds, D. L., Henderson, C. R., Jr., Hanks, C., Cole, R., Tatelbaum, R., McConnochie, K. M., Sidora, K., Luckey, D. W., Shaver, D., Engelhardt, K., James, D., & Barnard, K. (1997). Effect of prenatal and infancy home visitation by nurses on pregnancy outcomes, childhood injuries, and repeated childbearing. *Journal of the American Medical Association*, 278, 644–652.
- Layzer, J. I., St. Pierre, R., & Wilwerding, J. (1998, July). *Stability and change in the parenting skills and economic well-being of low-income mothers*. Paper presented at the Head Start Fourth National Research Conference, Washington, DC.
- Love, J. M., Nauta, M. J., Coelen, C. G., Hewett, K., & Ruopp, R. R. (1976, March). *National Home Start Evaluation: Final report*. Ypsilanti, MI: High/Scope Educational Research Foundation; Cambridge, MA: Abt Associates.
- Meisels, S. J., & Shonkoff, J. P. (Eds.). (1990). *Handbook of early childhood intervention*. New York: Cambridge University Press.
- Olds, D. L., Eckenrode, J., Henderson, C. R., Jr., Kitman, H., Powers, J., Cole, R., Sidora, K., Morris, P., Pettitt, L. M., & Luckey, D. (1997). Long-term effects of home visitation on maternal life course and child abuse and neglect: Fifteen-year follow-up of a randomized trial. *Journal of the American Medical Association*, 278, 637–643.
- Olds, D. L., Henderson, C. R., Jr., Chamberlain, R., & Tatelbaum, R. (1986). Preventing child abuse and neglect: A randomized trial of nurse home visitation. *Pediatrics*, 78, 65–78.
- Olds, D. L., & Kitman, H. (1993). Review of research on home visiting for pregnant women and parents of young children. *The Future of Children*, 3(3), 53–92.
- Pires, S. A. (Ed.). (1997). *Lessons learned from the Fort Bragg demonstration* (Louis de la Parte Institute Monograph No. CFS 155). Tampa: University of South Florida, Research and Training Center for Children's Mental Health.
- Prien, R. F., & Robinson, D. S. (Eds.) (1994). *Clinical evaluation of psychotropic drugs: Principles and guidelines in association with the NIMH and the ACNP*. New York: Raven Press.
- Provence, S., & Naylor, A. (1983). *Working with disadvantaged parents and their children: Scientific and practice issues*. New Haven, CT: Yale University Press.
- Ripple, C. H., Gilliam, W. S., Chanana, N., & Zigler, E. (1999). Will 50 cooks spoil the broth? The debate over entrusting Head Start to the states. *American Psychologist*, 54, 327–343.
- Sameroff, A. J., Seifer, R., Barocas, B., Zax, M., & Greenspan, S. (1987). IQ scores of 4-year-old children: Social-environmental risk factors. *Pediatrics*, 79, 343–350.
- Samuelson, R. J. (1998, February 23). Investing in our children: Sorry. Government programs can't undo most of the ill effects of family breakdown. *Newsweek*, p. 45.
- Seitz, V., & Apfel, N. H. (1999). Effective interventions for adolescent mothers. *Clinical Psychology: Science and Practice*, 6, 50–66.
- Seitz, V., & Provence, S. (1990). Caregiver-focused models of early intervention. In S. J. Meisels & J. P. Shonkoff (Eds.), *Handbook of early intervention* (pp. 400–427). New York: Cambridge University Press.
- Seitz, V., Rosenbaum, L. K., & Apfel, N. H. (1985). Effects of family support intervention: A ten-year follow-up. *Child Development*, 56, 376–391.
- Smith, S. (Ed.). (1995). *Two generation programs for families in poverty: A new intervention strategy*. New Jersey: Ablex Publishing Corporation.
- Striefel, S., & Robinson, M. A. (1998, February). *Lessons learned at the Community-Family Partnership Program: A Head Start Comprehensive Child Development Program*. Logan: Utah State University, Center for Persons with Disabilities.
- St. Pierre, R. G. (1998, March 26). Prepared testimony of Robert G. St. Pierre, Vice President, Abt Associates Inc. Presented at the Joint Hearing of the U.S. Senate and House Subcommittees

- (Senate Hearing 105–553). [Available on-line from CIS Congressional Universe (<http://web.lexis-nexis.com/congcomp/form/cong/pubs.html>) CIS-No. 98-S541-35].
- St. Pierre, R. G., & Layzer, J. I. (1998). Improving the life chances of children in poverty: Assumptions and what we have learned. *Social Policy Report: Society for Research in Child Development*, *XII*(4), 1–25.
- Valentine, J. (1997). Program development in Head Start: A multifaceted approach to meeting the needs of families and children. In E. Zigler & J. Valentine (Eds.), *Project Head Start: A legacy of the War on Poverty* (pp. 349–366). Alexandria, VA: National Head Start Association.
- Wagner, M. (1998). *The Parents As Teachers Program: Results from two demonstrations*. Menlo Park, CA: SRI International.
- Winter, M. M., & McDonald, D. S. (1997). Parents as teachers: Investing in good beginnings for children. In G. W. Albee & T. P. Gullotta (Eds.), *Primary prevention works. Issues in children's and families' lives* (Vol. 6, pp. 119–145). Thousand Oaks, CA: Sage.
- Yoshikawa, H. (1995). Long-term effects of early childhood programs on social outcomes and delinquency. *The Future of Children*, *5*(3), 51–75.
- Zigler, E., & Anderson, K. (1997). An idea whose time had come: The intellectual and political climate. In E. Zigler & J. Valentine (Eds.), *Project Head Start: A legacy of the War on Poverty* (pp. 3–20). Alexandria, VA: National Head Start Association.
- Zigler, E. F., Finn-Stevenson, M., & Stern, B. M. (1997). Supporting children and families in the schools: The School of the 21st Century. *American Journal of Orthopsychiatry*, *67*, 396–407.