DOCUMENT RESUME

ED 225 652 PS 013 297

AUTHOR	Affholter, Dennis; And Others
TITLE	The Infant-Toddler Component and Child Impact. Evaluation of the Child and Family Resource Program (CFRP).
INSTITUTION	Abt Associates, Inc., Cambridge, Mass.
SPONS AGENCY	Administration for Children, Youth, and Families (DHHS), Washington, D.C.
REPORT NO	AAI-80-75
PUB DATE	9 Dec 80
CONTRACT	HEW-105-79-1301
NOTE	160p.; For related documents, see ED 224 570, ED 224 611-614, and PS 013 299-303.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC07 Plus Postage.
DESCRIPTORS	*Child Development; Early Childhood Education; Family Involvement; *Family Programs; *Infants; Longitudinal Studies; *Low Income Groups; Outcomes of Education; Participant Characteristics; *Program Effectiveness; Program Evaluation
IDENTIFIERS	*Child and Family Resource Program; Developmental Continuity; Impact Studies; Project Head Start; *Toddlers

ABSTRACT

Fifth in a series of the Child and Family Resource Program (CFRP) evaluation reports, this volume focuses on the infant/toddler component of CFRP and its impact on children approximately a year and a half after they enter the program. A brief summary of the CFRP evaluation design and preliminary findings presented in previous evaluation reports is given in chapter 1. Chapter 2 provides a description of the infant/toddler component of the 11 CFRP's in terms of center-based activities, frequency and level of participation, and approaches used in both parent education and infant/toddler sessions. Chapter 3 examines CFRP impact on the development of the infants and toddlers involved; differences in means between CFRP children and those in the control/comparison group on the Bayley Scales of Infant Development (BSID) are reported in the attempt to identify major program impact. An executive summary of the study's findings is presented in chapter 4. Four appendices, accounting for approximately one half of the report, are also included: the first reviews issues related to the quality of the child assessment data reported, and the second discusses the rationale for the statistical tests chosen. A step-by-step description of the analysis of program impact reported in chapter 3 is presented in the third appendix, while the fourth describes analysis of the relationship between program participation and BSID scores. (MP)

ABT ASSOCIATES INC.
55 WHEELER STREET, CAMBRIDGE, MASSACHUSETTS 02138
TELEPHONE • AREA 617-492-7100

AAI No. 80-75

Project Director
Marrit J. Nauta

Project Officer
Esther Kresh, Ph.D.

EVALUATION OF THE CHILD AND FAMILY RESOURCE PROGRAM (CFRP)

The Infant-Toddler Component and Child Impact

Principal Authors
Dennis Affholter
Marrit J. Nauta

with
Lorelei Brush, Ph.D.
David Connell, Ph.D.

December 9, 1980

Editor
Lynell Johnson, Ph.D.

HEW-105-79-1301

Prepared for:

The Administration for Children, Youth and Families
Office of Human Development Services
Department of Health and Human Services
Washington, D.C. 20201

Contract Manager	Quality Control Reviewer	Management Reviewer

Table of Contents

Table of Contents (continued)

Tables and Figures

Page

FOREWORD

In 1973, the Administration for Children, Youth and Families (ACYF) initiated the Child and Family Resource Program (CFRP) as part of the Head Start Improvement and Innovation planning effort. CFRP was funded as a demonstration program with the intent of developing models for providing services to low-income families with young children--models which could be adapted by different communities serving different populations. There are eleven CFR programs across the country, one in each of the ten HHS (formerly HEW) regions and one representing the Indian and Migrant Division. Each program receives approximately $155,000-$170,000 per year to serve a minimum of 80 families.

CFRP is a family-oriented child development program which provides support services crucial for the sustained healthy growth and development of families who have children from the prenatal period through age eight. It promotes child development and meets children's needs by working through the family as a unit and provides continuity in serving children during the major stages of their early development. CFRP services are offered within the context of three major program components--infant-toddler, Head Start, and preschool-school linkage. Each is intended to serve families with children in a specific age group; all three taken together are intended to provide continuity--especially developmental and educational continuity--across the period of a child's life from before birth to the primary grades in school.

Another distinctive feature of CFRP is its emphasis on a comprehensive assessment of each family's strengths and needs and the development with the family of an individualized plan for services to be obtained through CFRP. Families enrolled in CFRP receive the same comprehensive services that are offered by Head Start and additional services tailored to the needs of each family. At the same time, CFRP works to reduce fragmentation and gaps in the delivery of services by existing community programs and agencies.

In October 1977, the Administration for Children, Youth and Families funded a longitudinal evaluation to determine the effectiveness of the Child and Family Resource Program. The evaluation is designed to address three major questions:

- What is the nature and extent of services that should be provided to families and children in order to meet their needs, enhance their strengths and foster independence?
- What are effective processes for the provision of these services?
- What can be learned about the developmental processes of families and how they relate to the developmental processes of children?

The current evaluation of CFRP was preceded by two other studies of the program, both also funded by ACYF. The first, conducted by Huron Institute in 1974-75, was an effort to determine the feasibility of a summative evaluation of CFRP. A formative evaluation of CFRP was also undertaken in 1974-75, by Development Associates Inc. A follow-up study was conducted by the same contractor in 1975-77.

This is the fifth in a series of CFRP evaluation reports. The first report presented the overall study design. Study implementation and the collection of baseline data on evaluation families were the focus of the second report. The third report consisted of three volumes: Volume I documented the first six months of the study and examined initial program impact on families; descriptive information about CFRP operations at the six evaluation sites was presented in Volume II; the third volume was a summary of the findings presented in the first two. The fourth report presented descriptive profiles of all eleven CFRPs.

This fifth report focuses on the infant-toddler component of CFRP and its impact on children approximately a year to a year and a half after they entered the program. Chapter 1 briefly summarizes the CFRP evaluation design and preliminary findings that were presented in previous evaluation reports. A description of the infant-toddler component in the eleven CFRPs is presented in Chapter 2. To the extent possible, we have attempted to identify program models for the delivery of services to families with children in the infant-toddler age range. We also report on the frequency of family participation in various infant-toddler activities at five of the six impact study sites through March 1980. These data are presented in an attempt to determine the extent to which the infant-toddler component as now operated is conducive to achieving the objectives of CFRP as articulated in the national Guidelines.

Chapter 3 examines CFRP impact on the development of the infants and toddlers who are the focus of this longitudinal evaluation. Differences in means between CFRP children and those in the control/comparison group on the Bayley Scales of Infant Development are tested in an attempt to identify any major program impact. In this chapter, we also explore the relationship between level of participation in CFRP and positive outcomes for children. An executive summary of findings is presented in Chapter 4.

There are four appendices to this report. Appendix A reviews issues related to the quality of the child assessment data reported here. Appendix B contains a discussion of analytic models often used in testing program impacts, and the rationale for the statistical tests chosen. Appendix C provides a step-by-step description of the analyses of program impact reported in Chapter 3. A description of the analyses of the relationship between program participation and Bayley scores is presented in Appendix D.

Acknowledgments

This report could not have been completed without the cooperation and assistance of numerous persons and groups. Several of these deserve special recognition for their contributions to this report and the evaluation effort.

We are especially grateful to Dr. Esther Kresh, the ACYF Project Officer for this evaluation, for her continuing guidance, assistance, and support. We also want to express our thanks to Martella Pollard, Director of the CFRP Demonstration, and to Ray Collins, Chief of the Development and Planning Division at ACYF, for their interest, enthusiasm, and guidance.

We wish to acknowledge the valuable assistance directors and staff at the six impact study sites have provided in the evaluation effort. They have spent numerous hours completing records and responding to questions about the operations of their programs. We also wish to thank the directors at the five non-impact study sites for providing us with an opportunity to learn about their programs.

The families in both the CFRP and control/comparison group at each of the six evaluation sites deserve special recognition. They permitted our staff to conduct developmental assessments on the infants and toddlers who are the focus of this evaluation.

We also wish to thank the National Advisory Panel for their guidance and assistance--Drs. Walter Allen, Tony Bryk, Jessica Daniel, Frank DiVesta, and Luis Laosa. In addition, we were fortunate to have Drs. (Ruth) Ann O'Keefe and Jean Carew as ad hoc members of our panel.

Finally, we would like to acknowledge the work of Abt Associates Inc. staff and consultants who played major roles in the preparation of this report. Special thanks go to Ruth Wolman, a consultant, who directed the Bayley training sessions for trainers and on-site data collection staff and provided ongoing monitoring of data quality. The management of the data collection effort was anchored by Ilona Ferraro, Jan Stepto-Millett, and Lucy Algere-Knox. We also wish to acknowledge the special role of our on-site staff at the six impact study sites who had responsibility for the tracking of families and for the child development assessments. The analytic effort was skillfully managed by Dennis Affholter, Lorie Brush, and David Connell, with the support of Lucy Algere-Knox and Roz Ladner. Lynell Johnson provided editorial support for this report. Finally, special thanks go to our administrative and secretarial staff for the numerous ways in which they assisted project staff--Patricia McMillan, Annie Hondrogen, and Kathe Phinney.

Marrit J. Nauta
Project Director

Chapter 1

EVALUATION OVERVIEW

The CFRP evaluation, funded in October 1977 by the Administration for Children, Youth and Families (ACYF), seeks to provide detailed information about the effectiveness of CFRP as a whole, of individual programs, and of particular program elements or configurations of elements. Such information can aid ACYF in making decisions about expansion of the program and/or dissemination of its most important and effective features.

The first phase of the CFRP evaluation was devoted to redesign, start-up of the study, and collection of baseline data. In Phase II, the evaluation examined program impact on families after six months in the program, as well as CFRP treatment and processes used to deliver services to families. The major focus of the third phase of the evaluation is on program impact after families have participated in CFRP for a year and a half.

The initial design for the CFRP evaluation consisted of three distinct but interrelated components which address the following objectives:

- description of CFRPs and their operations;
- identification of program models;
- linking of family outcomes to particular aspects of CFRP treatment (characteristics of staff and program) and to family characteristics; and
- linking of family outcomes to participation or nonparticipation in CFRP.

The three component studies--program, impact, and process/treatment--are complementary ways of viewing the effects and effectiveness of CFRP.

The program study is designed for the purpose of developing a comprehensive picture of the operations of CFR programs. Information collected during site visits and in interviews with program staff is used to develop profiles of program implementation and to establish a descriptive context for the statistical and analytic findings of other components of the evaluation. Site visits took place at six of the eleven CFRPs in fall 1978, spring 1979, and spring 1980: Jackson, MI; Las Vegas, NV; New Haven, CT; Oklahoma City, OK; St. Petersburg, FL; and Salem, OR. These six programs were not randomly selected; they were chosen on the basis of their ability to recruit the requisite number of families for the impact study. Brief interviews were conducted in spring 1980 with staff from the five non-impact study CFRPs to obtain descriptive information about the operations of these programs.

The impact study is designed to determine the effects of CFRP services on families by comparing CFRP families with a group not enrolled in the program. At the six sites listed above, families entered the evaluation when they had a child less than one year old and were randomly assigned either to CFRP or to a control/comparison group. At entry into the evaluation, there were an average of 39 CFRP and 38 control/comparison families per site. These families will be followed until the focal child has completed at least one year of elementary school (1985).

The impact study focuses on five outcome domains likely to be affected by family participation in CFRP:

- family circumstances (e.g., employment, education);
- maternal and child health;
- parent-child relationship and interaction;
- child development and achievement; and
- capacity for independence (use of community resources, locus of control and coping strategies, affiliation with family and social networks).

This report focuses on the domain of child development and achievement. Data concerning child development were obtained for the first time in fall/winter of 1979-80. Program impact on families in the other four outcome domains was examined in the Phase II Report. There was little evidence that the program had had a positive impact on the sample families in these domains after six months of participation. This may be due partly to the fact that such a period is too short for impact to become apparent. For example, changes in family circumstances or capacity for independence may not become evident until the family has been involved in the program for a longer period of time. However, results of a pilot study of parent-child interaction conducted in spring 1979 did provide preliminary evidence of program impact in this area. Specifically, CFRP mothers had more frequent interactions with their children than was the case for mothers in the control/comparison group--although these findings were largely site-specific.* Positive changes in parent-child interaction are expected to influence the development of the child.

To date, impact study data have been obtained at four time points: fall 1978 (baseline), spring 1979 (six months after the families entered the evaluation), fall/winter 1979-80 (after approximately one year of program participation),

*Results of the parent-child interaction observation pilot study are reported in Appendix E of the <u>Phase II Report, Volume I: Research Report</u>, February 25, 1980.

and spring 1980 (after about 18 months of treatment). Sample sizes at the first three data collection points (not including spring 1980) are noted in Table 1-1. As is evident from the table, sample attrition was most severe during the first six months of the CFRP evaluation. By fall/winter 1979-80 the attrition rate appeared to have leveled off considerably. This was partly due to a decision to retain families in the evaluation who decided they no longer wanted to participate in CFRP. There were 13 former CFRP families in the fall/winter sample (8 in Jackson and 5 in St. Petersburg).

Table 1-1
Impact Study Sample Sizes

	Fall 1978		Spring 1979		Fall/Winter 1979-80	
	CFRP	Control	CFRP	Control	CFRP	Control
Jackson	40	24	31	20	30	20
Las Vegas	42	43	32	33	35	29
New Haven	36	20	28	18	26	14
Oklahoma City	39	49	32	45	28	43
St. Petersburg	40	43	34	40	34	38
Salem	39	51	31	42	34	40
Total	236	230	188	198	187	184
Site Average	39	38	31	33	31	31
Attrition			21%	13%	-	6%

The process/treatment study focuses on the CFRP families who participate in the impact study at the six sites. This study is designed to explore relationships among characteristics of families and staff, interactions between staff and families, services provided, and program impact. Data were collected in fall 1978, spring 1979, and spring 1980 through interviews with staff and families. In addition, ongoing data collection systems are being maintained for data concerning family participation in the program, family goals, and referrals for services.

A fourth component has been added to the evaluation in Phase III. The <u>ethnographic study</u> is designed to broaden our understanding of how CFRP works with families and functions as a child development and family support program. Data will be gathered through a series of in-depth interviews and observations to be conducted beginning in fall 1980.

Chapter 2

THE INFANT-TODDLER COMPONENT

> The primary objective of the infant-toddler component of CFRP is "to assist parents to promote the total (emotional, cognitive, language, and physical) development of infants and toddlers through age three." The eleven CFRPs attempt to achieve this objective through home visits and center-based activities--parent education sessions and infant-toddler sessions. However, at most sites, parent education sessions offer little or no direct demonstration of techniques for working with children, and home visits tend to focus on helping parents meet specific needs rather than on enhancing the development of the child. Moreover, levels of participation in both center-based and home-based program activities are low in many instances. Thus there is reason to doubt that the infant-toddler component, as currently implemented, can have significant positive effects on children's development.

According to the national CFRP Guidelines, the objective of CFRP's infant-toddler component is to enhance the total development of infants and toddlers through age three. There are essentially three different approaches that could be used to attain this goal: (1) direct intervention with children; (2) parent education to assist parents in their role as primary educators of their own children; and (3) a combination of the two. The second approach is advocated in the Guidelines; it is by working through parents and the family as a unit that CFRP expects to influence the development of children. Numerous research studies support this focus. The evidence indicates that parent education not only can be an effective strategy in promoting child development, but may be a necessary step if any lasting improvement in the child's functioning is to be attained. Thus, involvement of the child's parents as active participants appears to be critical to the success of a child development program such as CFRP.

In practice, the parent education approach to providing infant-toddler services has been adopted generally by local CFRPs, with some secondary emphasis on direct intervention. Parent education sessions and home visits are the principal parent-focused activities at the eleven CFRPs. In addition, most of the programs offer some form of group activity for children ("infant-toddler sessions").

This chapter presents descriptive information on the infant-toddler component of CFRP. Section 2.1 deals with center-based activities, including frequencies, levels of participation, and approaches used in both parent education sessions and infant-toddler sessions. Section 2.2 presents parallel information for home visits. Section 2.3 summarizes data on participation for all program activities combined.

2.1 Center-Based Activities

2.1.1 Frequencies and Participation Rates

While most CFRPs offer two to four center sessions per month (Table 2-1), most parents actually attend much less frequently. Family participation in center sessions is viewed by all programs as "less than optimal." For five of the six evaluation programs*, this means that only 39 percent of the study families were involved in center sessions an average of once or more per quarter since they enrolled in CFRP. Participation varied from site to site, however, as noted in Table 2-2. Participation was particularly problematic in Oklahoma City and Las Vegas. Problems with attendance in Oklahoma City

*The sixth, New Haven, was excluded from participation analyses due to a high incidence of missing data.

Table 2-1

Frequency of Center-Based Activities

	2 times/ week	Weekly	3 times/ month	2 times/ month	Monthly
Bismarck		x			
Gering			x		
Jackson				x	
Las Vegas				x	
Modesto		x[a]		x[a]	
New Haven	x				
Oklahoma City				x	
Poughkeepsie				x	
St. Petersburg					x[b]
Salem		x			
Schuylkill Haven		x			

[a]Frequency of parent education sessions varies depending on the season. At harvest time, sessions occur twice a month; weekly sessions are offered in the off season.

[b]St. Petersburg's monthly center sessions are supplemented by weekly study groups.

Table 2-2

Participation in Center Sessions[a]
(percent of families)

	Jackson	Las Vegas	Oklahoma City	St. Petersburg	Salem	Overall
N.	38	36	30	34	39	177
At least once per quarter	61	17	13	50	49	39
Less than once per quarter	39	83	87	50	51	61

[a]Figures are based on the ongoing record-keeping system for CFRP families in the impact study. (See chapter note for an explanation of techniques used to compute participation rates.)

undoubtedly were due to the fact that center sessions were not offered for some time during the first year and a half after the families entered CFRP; center-based activities were resumed at this site in winter of 1980. In Las Vegas, problems with center attendance can be attributed to characteristics of the families that were selected for the evaluation; many of the teenage mothers in the evaluation sample attend school during the day and are unable to participate in daytime center activities. Center participation was less problematic in Jackson, St. Petersburg, and Salem, where 49 to 61 percent of the families attended sessions regularly.

Table 2-3 shows participation in center sessions for only those evaluation families who attended regularly, i.e. at least once per quarter on average (henceforth termed "center" families). Most of these families (68%) attended one to three sessions per quarter; very few participated in all sessions that were offered by their local programs. Families who came to the center regularly attended an average of 3.4 sessions per quarter. Mean attendance rates ranged from a low of 2.2 sessions in Las Vegas to a high of 4.3 in St. Petersburg. (St. Petersburg rates are high partly because the data include weekly study groups as well as monthly parent education sessions.) Participation of other families in the evaluation sample (those attending less than once per quarter) averaged .24 sessions per quarter, or one center session every 12 months.

Tables 2-4 to 2-7 break out participation rates for infant-toddler and parent education sessions, respectively, for the sample of families whose children were tested in connection with the child impact study. (In Chapter 3, children's test performance is examined in relation to attendance at these two types of center sessions, taken separately.) The data for this somewhat different sample (see table notes for explanations of the differences) are largely consistent

Table 2-3

Center Participation per Quarter[a]
(percent of "center" families)

	Jackson	Las Vegas	Okla-homa City	St. Peters-burg[b]	Salem	Overall
Number of sessions offered per month	2	2	2	1	4	--
N of families	23	6	4	17	19	69
Number of sessions attended per quarter						
1	35	33	75	6	21	26
2	26	50	0	35	16	26
3	17	17	0	18	16	16
4 or more	21	0	25	42	47	32
Mean number of sessions attended per quarter	2.8	2.2	3.2	4.3	3.8	3.4

---

[a]See chapter note for an explanation of computational procedures used to derive these percentages. New Haven is omitted from the table because of its high incidence of missing data. Sample includes all families who participated in center sessions at least once per quarter and who remain active participants in CFRP.

[b]Data for St. Petersburg include weekly study groups as well as monthly parent education sessions.

Table 2-4

Participation in Parent Education Sessions
(percent of families)[a]

	Jackson	Las Vegas	Okla-homa City	St. Peters-burg	Salem	Overall
N	24	37	27	25	28	141
Less than once per quarter	33	92	93	56	54	68
Once or more per quarter	67	8	8	44	46	32

---

[a]Sample includes families whose children were tested in the child impact study. Some of these families have since terminated participation in CFRP.

Table 2-5

Participation per Quarter in Parent Education Sessions (percent of "center" families)[a]

	Jackson	Las Vegas	Oklahoma City	St. Petersburg	Salem	Overall
N	16	3	2	11	13	45
Number of sessions attended per quarter						
1	31	100	50	55	38	44
2	25	0	0	27	31	24
3	19	0	0	0	15	11
4 or more	25	0	50	18	15	20
Mean number of sessions attended per quarter	2.9	1.0	3.3	2.3	2.5	2.6

[a]Sample includes families whose children were tested in the child impact study. Some of these families have since terminated participation in CFRP.

Table 2-6

Participation in Infant-Toddler Sessions (percent of families)[a]

	Jackson	Las Vegas	Oklahoma City	St. Petersburg	Salem	Overall
N	24	37	27	25	28	141
Less than once per quarter	33	84	96	56	50	66
Once or more per quarter	67	16	4	44	50	34

[a]Sample includes families whose children were tested in the child impact study. Some of these families have since terminated participation in CFRP.

Table 2-7

Participation per Quarter in Infant-Toddler Sessions (percent of "center" families)[a]

	Jackson	Las Vegas	Oklahoma City	St. Petersburg	Salem	Overall
N	16	6	1	11	14	48
Number of sessions attended per quarter						
1	31	100	0	27	50	44
2	19	0	0	36	29	23
3	25	0	100	0	21	17
4 or more	25	0	0	36	0	17
Mean number of sessions attended per quarter	3.1	1.0	3.0	2.8	2.2	2.3

[a]Sample includes families whose children were tested in the child impact study. Some of these families have since terminated participation in CFRP.

with the data in Table 2-3: For each type of center session, there are a significant number of nonparticipating families, especially in Las Vegas and Oklahoma City. Among those who do participate, frequencies are typically in the range of one to three times per quarter, although there are a few families at certain sites who participate much more frequently.

CFRP staff attribute occasional nonparticipation mostly to illness, crises, or emergencies that prevent parents from attending. However, chronic nonparticipation on the part of some families represents a problem for all programs. Some mothers simply do not wish to join a group or do not believe they will benefit from being involved. Others consider it "too risky" to attend--in the sense of feeling vulnerable or deficient--or lack the necessary support from husband or family.

As already implied, participation is particularly problematic for mothers who are employed or attend school during the day, when center sessions typically take place. Gering and St. Petersburg are the only programs that conduct evening sessions for parents on a regular basis--once a month--to accommodate the working or in-school mother. In other programs, evening sessions are a rare occurrence. Some programs that have tried them found that participation did not increase; parents are simply too tired after a full day of work or school to attend, or do not want to take even more time away from being with their children.

A variety of approaches are used by local programs in an attempt to increase participation in center sessions. All CFRPs except one provide transportation for parents who could otherwise not attend. Several programs hold their center sessions in more than one location to make them more accessible. Others offer some sort of tangible incentive or have established policies concerning minimum participation in center-based activities; these policies appear to have a positive influence on attendance rates.

2.1.2 Center Session Content and Models

Parent education sessions are intended to provide families with a basic knowledge of child growth and development and to assist them in developing more effective parenting skills. Infant-toddler sessions are intended to provide children with a group experience and give them an opportunity to learn to share and get along with others. In addition, some programs emphasize acquisition of skills, such as language, cognitive, motor, social-emotional, and self-help.*

*In a few CFRPs, sessions for infants and toddlers are guided by the same curriculum as that used for home visits. It is not uncommon for home visit staff to participate in these sessions. Each program has developed special mechanisms to ensure some level of continuity between center- and home-based activities, either in the form of records or periodic meetings with appropriate staff, although the degree of coordination is not high in most cases. (See the next section for discussion of the content of the home visit curriculum.

All are designed to enhance the child's overall development, and these approaches prepare the child for entry into Head Start. (Individualization of activities to meet the specific needs of each child is quite limited in most programs.)

Despite these indications of a concern for the education and development of the child, however, center-based sessions at most sites are not organized in a fashion that is likely to maximize their developmental effects. Two models for integrating the parent education and infant-toddler portions of center-based sessions are currently in operation at the eleven CFRPs:

- The Parent-Child Interaction Model provides extensive opportunity for involvement of parents with their own children at the center. Center sessions are designed to help parents acquire effective child care techniques and to teach them developmental activities that are appropriate to the child's needs. Classroom staff assist parents in this task and provide feedback on parent-child interactions. The group discussions that follow focus on topics related to child development or child-rearing practices.

- In the Separate Parent-Child Session Model, parent education focuses almost entirely on parents, away from their children. Children are cared for in an infant-toddler room while parents attend parent education sessions. There is little or no opportunity for parents to interact with their children at the center.

The second model is likely to be somewhat less effective, because it relies mostly on lectures and other didactic approaches as methods of parent training. There is little evidence that simply providing information to parents will in itself lead to significant change in parental behavior or skills.* Observation of modeled behavior, which is regarded

*Bronfenbrenner, U. Is Early Education Effective? Washington, D.C. DHEW Publication No. OHD 74-75, 1974.

as a more effective learning tool, is used extensively in the Parent-Child Interaction Model. This model is in place at only three programs--Bismarck, Gering, and New Haven; all other programs conduct separate sessions for parent and child. (Salem is a partial exception, in that it has offered opportunities for parent-child interaction to selected families with toddlers who have special needs; its regular parent education program involves separate sessions.)

2.2 Home Visits

2.2.1 Scheduled and Actual Frequencies

Center sessions are not the only mechanism for educating parents of infants and toddlers. A regular home visiting program can also help parents to strengthen their child-rearing skills and increase their knowledge about child development. Home visits are an integral part of the infant-toddler component at all eleven programs. The potential importance of the home visits is underscored by indications, discussed in the previous section, that center-based parent education may be less than optimal as currently conducted.

There is some question, however, whether visits occur with sufficient frequency to carry out an effective parent education program in the home. (The importance of frequent visits was shown in the evaluation of the Home Start Demonstration Program, which found a strong relationship between visit frequency and school readiness and language development scores of preschool children.*) Scheduled frequencies of CFRP home visits range from one to four times a month (Table 2-8). However, it is evident from program records and discussions with CFRP staff that at some sites, at least, home visits occur less frequently than the schedule called for in local program plans (Table 2-9).

*Love, J.M., Nauta, M.J., Coelen, C.G., et al. National Home Start Evaluation: Final Report--Findings and Implications, High/Scope Educational Research Foundation, Michigan, and Abt Associates Inc., Massachusetts, 1976.

Table 2-8

Home Visit Frequency

	3 times/ month	2 times/ month	Monthly	Varied[a]
Bismarck				x
Gering		x		
Jackson	x			
Las Vegas		x		
Modesto		x		
New Haven				x
Oklahoma City			x	
Poughkeepsie		x		
St. Petersburg				x
Salem				x
Schuylkill Haven			x	

[a]In these programs, home visit frequency varies depending on family need and interest.

In addition to showing the relationship between scheduled and actual home visits, Table 2-9 illustrates another important point--namely, that home visits and center attendance go hand in hand, rather than being alternative or complementary ways in which families take part in CFRP. Families who participate in center sessions less than once per quarter ("non-center" families) receive considerably fewer home visits than families who come to the center regularly. Families in the latter group ("center" families) were visited about two times per month on the average, while "non-center" families were seen only once a month. Only in Las Vegas were the two groups of families involved in home visits at approximately the same rate. Across all sites, the correlation between home visit rate and center participation was .49.

These differences in home visiting rates for the two groups are somewhat surprising. One might have expected home visits to occur with greater intensity with families who

Table 2-9

Home Visit Participation per Quarter

(percent of families)*

	Jackson Center	Jackson Non-Center	Las Vegas Center	Las Vegas Non-Center	Oklahoma City Center	Oklahoma City Non-Center	St. Petersburg Center	St. Petersburg Non-Center	Salem Center	Salem Non-Center	Overall Center	Overall Non-Center
Home visits offered per month	3		2		1		varied		varied		---	
N of families	23	15	6	30	4	26	17	15	19	20	69	108
Number of home visits per quarter												
1	0	20	67	30	25	73	17	27	5	10	7	36
2	0	13	17	47	50	15	12	7	11	5	9	20
3	4	27	0	10	25	12**	18	13	16	40	9	19
4	13	7	17	13	0	0	6	33	16	25	14	14
5	4	20	0	0	0	0	18	7	26	0	7	4
6	17	0	0	0	0	0	12	7	16	5	17	2
7	17	7	0	0**	0	0	12	0	5	5	13	2
8	17	0	0	0	0	0	6	0	5	10	10	2
9	9	0**	0	0	0	0	6	7	0	0	6	1
10	9	7	0	0	0	0	0	0	0	0	4	1
11	4	0	0	0	0	0	0	0	0	0	1	0
12	4	0	0	0	0	0	0	0	0	0	1	0
Mean number of home visits per quarter	7.61	4.04	2.33	2.41	2.37	1.68	5.96	3.61	5.80	4.19	5.94	2.98
S.D.	2.36	2.51	1.15	1.01	.88	.70	2.38	2.01	1.77	1.89	2.63	1.86

*Center families are those who participated in center sessions at least once per quarter; non-center families attended less frequently.

**The underline denotes the number of home visits that are supposed to take place according to individual program schedules.

never come to the center or come only occasionally, in order to ensure that all families receive all of the benefits offered by CFRP. Instead, it appears that non-center families are simply less committed to CFRP than other families served by the program, perhaps due to a lack of interest or motivation to participate or, in the opinion of parents, less need for CFRP services.

2.2.2 Home Visit Content and Models

In general, home visits do not represent a continuation of the curriculum or activities presented at center-based parent education sessions. At most sites, there is no explicit attempt to follow up on center activities in the home. One reason is that, while an effort is made to adapt center sessions to the needs of those present, they are nevertheless group sessions. Home visits, on the other hand, can be highly individualized.

Discussions with staff suggest that child-related parent education activities provided in the home typically involve helping parents to use elements in the child's environment as teaching tools and to turn everyday experiences into constructive learning situations. Parents are reminded about the teaching potential of all household tasks and the many objects in the home that can be used as instructional materials. In some programs, staff bring specific activities into the home to involve both parent and child. Usually the activity is preceded by an explanation of its importance and how it fits into the overall development of the child. An attempt is made not only to demonstrate activities to the parent, but to get her actively involved in working with the child. Frequently, a different set of activities is selected or planned for each family to ensure that they are appropriate to meet specific parent or child needs.

Despite the potential value of these activities for child development, it is not clear that they receive adequate emphasis in CFRP's home visits. In most programs, home visits are designed to have a dual focus: (1) helping parents to become more effective in their role as educators of their children; and (2) helping parents to meet a broad range of family needs and concerns. There appear to be differences among the eleven programs in the relative emphasis that is placed on parent education and family needs. The evidence suggests that home visiting staff in some programs devote only minimal attention to parent education or child development concerns. This is not true in all programs, however. In fact, at two sites, the dual focus of home visits is explicitly recognized, and separate family workers are assigned the responsibility for each aspect. Two different models of infant-toddler home visit assignments are currently in place within local CFRPs:

- The Team Model--employed in Jackson and New Haven--was developed to ensure that both parent education concerns and family needs are addressed adequately in home visits. Visits are conducted by two family workers: one has responsibility for working with the parent and child on issues related to the child's development and parenting skills; the other focuses more broadly on family needs.

- The Single Worker Model--employed at all other sites--assigns one family worker to each family, with responsibility for both aspects of the home visits, child development and parenting issues as well as family needs.

Home visit emphasis is determined to some extent by the type of curriculum that is used to guide home visit activities. Only four programs have adopted a developmental curriculum. In the other seven programs, home visits are planned by family workers themselves. This effort is closely supervised in three of these programs, usually by someone with a background in child development; family workers in the other four programs receive little or no guidance concerning the types of activities that visits should cover. It is of interest to note that there appears to be a relationship between the home visit planning effort and the frequency with which home visits occur; frequency decreases when home visiting staff do their own planning and have no curriculum or supervisor to fall back on (Table 2-10).

The fact that greater emphasis appears to be placed on family needs in the home-based activities of some programs may be related to the background of family workers. Their training tends to be in social work or related fields, rather than in parent education or child development. A substantial proportion (56%) of family workers at the six impact study sites perceive a need for additional training in these areas.*

Table 2-10

Relationship between Home Visit Frequency and Curriculum[a]

	Home Visit Frequency: Once a month	Home Visit Frequency: More than once a month
No curriculum or specific help		
Oklahoma City	x	
Schuylkill Haven	x	
Curriculum		
Jackson		x
Las Vegas		x
Poughkeepsie		x
Specific help		
Gering		x
Modesto		x

[a]Programs with varied home visit frequency were excluded from these analyses.

*Phase II Program Study Report, February 1980.

There are factors other than those related to staff and program characteristics that influence home visit focus. The types of activities that take place are dictated largely by the needs and goals that individual families identify or set for themselves. It is up to parents to decide what they want out of the program, a philosophy that is inherent in the CFRP concept. As impact study data indicate, the majority of families appear to be more concerned with getting help with such practical needs as housing, employment, and health than with child development or parenting skills.* (This is the case at least in the early stages of families' involvement in CFRP; the types of goals that are set may change after families have been in the program for longer periods of time.) In some programs parents also have a choice in deciding what program activities they want to participate in and with what frequency, factors that undoubtedly contribute to low participation levels in center-based activities. Yet few programs have developed special home-based activities for families that do not attend center sessions to ensure that appropriate parent education services are provided in the home. While individualization (interpreted as parent choice) of program services is an explicit mandate of CFRP, it is not clear that it maximizes attainment of the child development-related objectives of the program.

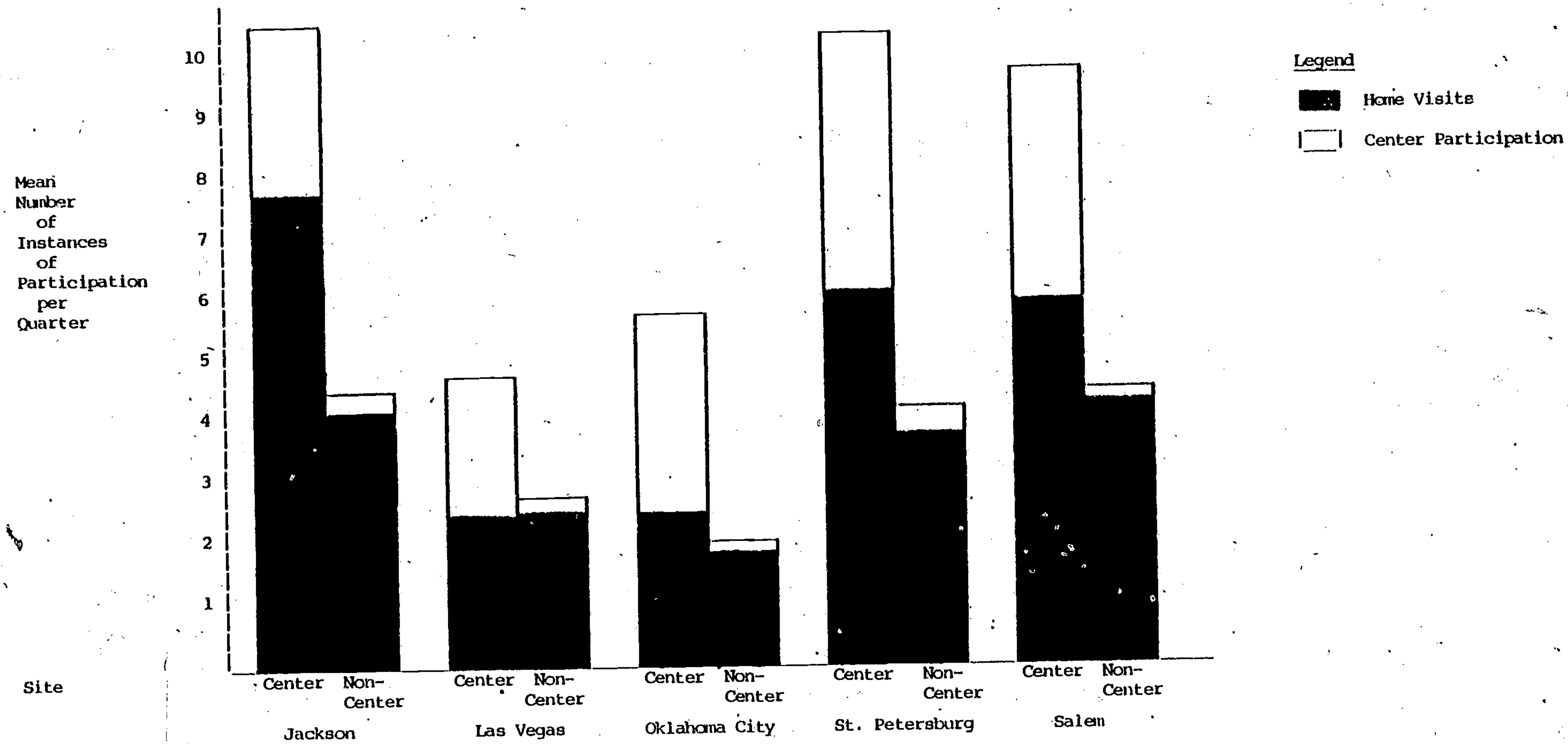
2.3 Total Program Participation

Figure 2-1 summarizes data on program participation by evaluation families presented earlier in this chapter. Shown are total participation rates (including center sessions and home visits) for both "center" families (those who came to the center at least once per quarter) and "non-center" families (those who participated less frequently in center-based activities.) Total participation is consistently lower

*Phase II Research Report, February 1980.

Figure 2-1

Participation in CFRP



for non-center families. As is evident from the figure, treatment is considerably more intensive in Jackson, St. Petersburg, and Salem than in Las Vegas and Oklahoma City.

The relatively low participation rates of "non-center" families, and of nearly all families at certain sites, raises serious questions as to whether CFRP can be expected to have an overall effect on children's development. While effects might be expected for active participants, any such effects might be diluted by nonparticipant families--and thus overall comparisons between program families and those outside the program might fail conventional statistical tests. As we have seen, this concern is underscored by two additional facts: (1) The content of home visits is not in all cases focused on child development; and (2) the methods by which center-based parent education sessions are conducted do not involve the kind of practical instruction that is likely to be most useful. The next chapter demonstrates that these concerns are justified, in that CFRP has negligible developmental effects, but also that there are promising signs that active participation may convey developmental benefits.

Chapter 2 Note

Participation rates reported in Tables 2-2 and 2-3 were computed as follows: Monthly counts of frequencies of participation were averaged to the quarterly level, and quarterly data were then averaged across the six quarters comprising the 18-month period to arrive at overall participation rates. In the event that data were missing for a quarter, rates were computed based only on quarters for which data were available. In cases where data were missing for one month of a quarter, data from the remaining two months were used to compute a quarterly participation rate. However, if data were missing for more than one month, data were considered missing for that entire quarter. These computational procedures differ from those used in earlier reports. Partly because of these changes in procedure, but particularly because most tables in this chapter include only "center" families who participated at least once per quarter, participation rates reported here are higher than those reported previously. It should also be noted that CFRP staff did not always distinguish between infant-toddler and parent education sessions when reporting participation rates, so that total center participation figures (Tables 2-2 and 2-3) are more meaningful than either parent education or infant-toddler rates reported separately.

Chapter 3

PROGRAM IMPACT ON CHILD DEVELOPMENT

> Scores of 187 CFRP and 184 control/comparison group children on the Bayley Scales of Infant Development were compared across and within the six impact study sites, employing both ANCOVA and value-added approaches. None of these analyses revealed any significant differences between the two groups, with the exception of a positive effect of CFRP on mental growth at one site, assessed by the value-added method. However, analyses of relationships between CFRP participation variables and Bayley scores within the CFRP group indicate that participation in infant-toddler sessions may be a key element in enhancing child development. Whether considered alone, in interaction with home visit rate, or in interaction with a measure of family strengths, rate of participation in infant-toddler sessions is positively associated with children's development.

This chapter examines CFRP's impact on the development of the infants and toddlers who are the focus of this longitudinal evaluation. Their development was measured by the Bayley Scales of Infant Development. Section 3.1 describes the sample of children tested and their performance on individual items in the Bayley Scales. Section 3.2 describes the analytic approach employed. Comparisons of performance between CFRP children and those in the control/comparison group on the Bayley Scales are reported in Section 3.3. Relationships between level of participation in CFRP and outcomes for children are discussed in Section 3.4. Finally, Section 3.5 presents a brief summary of the results.

3.1 Sample and Test Description

The Bayley Scales were administered to infants and toddlers at the six impact study sites over a period of six months (October 1979 to March 1980). Testing was staggered so that younger children would reach 15 months of age before being tested. The vast majority (95%) of children tested were between the ages of 15 and 22 months. Children were tested in their own homes, since a pilot study revealed that in-home testing could be feasible with only minor modifications of the instrument (described below). On-site researchers who underwent a one-week Bayley training session conducted these assessments; tester performance in Bayley administration was monitored throughout the data collection period. Overall, the quality of the assessment data was judged to be good. (Details of the monitoring process and analyses of data quality are reported in Appendix A.)

3.1.1 The Sample

Characteristics of the child sample were examined in order to determine whether there were differences from site to site, or differences between CFRP and control/comparison children within or across sites. The most important difference revealed by this examination had to do with child age. Ages of children at time of assessment were distributed differently across the six impact study sites (Table 3-1). Age differences also were evident, within and across sites, between children in the CFRP and control/comparison groups. Overall, 61 percent of the children tested were 18 months of age or younger; a higher proportion of the control/comparison children (66%) than of CFRP children (56%) were in this younger group. This difference in age distributions had major analytic consequences, as discussed later.

In addition, differences were detected on selected family background characteristics both across and within

Table 3-1

Ages of Children at Time of Testing

	15-16 months	17-18 months	19-20 months	21-22 months	23-24 months	Greater than 24 months	Total
Total sample	115(31%)	111(30%)	70(19%)	55(15%)	15(4%)	5(1%)	371(100%)
Jackson							
CFRP	6(20%)	15(50%)	5(17%)	2(7%)	2(7%)	0	30
Comparison	8(40%)	6(30%)	4(20%)	2(10%)	0	0	20
Las Vegas							
CFRP	13(37%)	9(26%)	6(17%)	6(17%)	1(3%)	0	35
Comparison	11(38%)	8(28%)	5(17%)	4(14%)	1(3%)	0	29
New Haven							
CFRP	5(19%)	3(12%)	10(38%)	5(19%)	2(8%)	1(4%)	26
Comparison	3(21%)	2(14%)	4(29%)	4(29%)	1(7%)	0	14
Oklahoma City							
CFRP	9(32%)	7(25%)	5(18%)	5(18%)	1(4%)	1(4%)	28
Comparison	9(21%)	18(42%)	7(16%)	6(14%)	2(4%)	1(2%)	43
St. Petersburg							
CFRP	10(29%)	8(24%)	7(21%)	7(21%)	1(3%)	1(3%)	34
Comparison	15(37%)	19(48%)	4(10%)	2(5%)	0	0	40
Salem							
CFRP	9(26%)	11(32%)	6(18%)	6(18%)	2(6%)	0	34
Comparison	17(45%)	5(13%)	7(18%)	6(16%)	2(5%)	1(3%)	38

sites. Several characteristics were examined: per capita income, sex of the focal child, employment and marital status of the mother, whether the mother had graduated from high school, and whether the focal child was the oldest. Site differences were evident for marital status and whether the tested child was the oldest in the family. Some differences also were detected between the CFRP and control/comparison groups at selected sites:

- child's sex--Las Vegas and Salem
- mother's employment status--Jackson, New Haven, Oklahoma City, St. Petersburg, and Salem
- marital status--New Haven
- whether the child is the oldest--New Haven, St. Petersburg, and perhaps Jackson

3.1.2 Overall Performance on Bayley Items

The Bayley instrument consists of two scales, a mental development scale (MDS) and a physical development scale (PDS). Items are normed by age. From the MDS a total of 69 items, normed for ages twelve to over thirty months, were selected for administration to the study's sample of infants and toddlers. Table 3-2 shows the percentage of children passing each item, by age. With a few exceptions, most MDS items were passed with greater frequency by older than by younger children. (All items were retained for further analysis, in order to preserve comparability with published norm data for the scale as a whole.)

Testers also administered 16 items from the PDS that required no special equipment (i.e., steps or walking board) and were therefore feasible to use in the home. Performance varied substantially on only a few items (Table 3-3). Ten items that had adequate variation in pass/fail rates (indicated with an asterisk in Table 3-3) were retained for future analysis.

Table 3-2

Percentage Passing MDS Items

Item	Normed age	15-16 months N=115	17-18 months N=111	19-21 months N=100	22-24 months N=40	All children N=366
101. Jabbers expressively	12.0	98	100	100	100	99
102. Uncovers blue box	12.0	98	100	100	100	99
103. Turns pages of book	12.0	100	100	100	100	100
104. Pats whistle doll, in imitation	12.2	98	99	100	100	98
105. Dangles ring by string	12.4	97	100	100	100	98
106. Imitates words	12.5	87	96	97	100	94
107. Puts beads in box (6 of 8)	12.9	97	100	99	100	99
108. Places 1 peg repeatedly	13.0	91	95	98	100	95
109. Removes pellet from bottle	13.4	90	96	98	100	95
110. Blue board: places 1 round block	13.6	72	86	94	93	85
111. Builds tower of 2 cubes	13.8	76	84	88	98	84
112. Spontaneous scribble	14.0	97	90	92	95	93
113. Says two words	14.2	60	66	79	90	70
114. Puts 9 cubes in cup	14.3	56	67	79	90	69
115. Closes round box	14.6	68	76	83	98	77
116. Uses gesture to make wants known	14.6	97	97	96	100	97
117. Shows shoes or other clothing or own toy	15.3	74	80	86	93	81
118. Pegs placed in 70 seconds	16.4	40	64	80	98	65
119. Builds tower in 3 cubes	16.7	27	40	62	83	47
120. Pink board: places round block	16.8	60	61	70	70	64
121. Blue board: places 2 round blocks	17.0	37	45	60	75	50
122. Attains toy with stick	17.0	54	64	78	83	67
123. Pegs placed in 42 seconds	17.6	12	36	66	87	43
124. Names 1 object	17.8	15	31	46	60	34
125. Imitates crayon stroke	17.8	38	48	48	65	47
126. Follows directions, doll	17.8	38	59	70	83	59
127. Uses words to make wants known	18.8	27	42	54	73	45
128. Points to parts of doll	19.1	12	13	48	68	29
129. Blue board: places 2 round and 2 square blocks	19.3	11	31	42	70	33
130. Names 1 picture	19.3	10	24	43	58	30
131. Finds 2 objects	19.7	30	29	31	50	32
132. Points to 3 pictures	19.7	1	10	30	40	17
133. Broken doll: mends marginally	19.9	16	24	32	48	27
134. Pegs placed in 30 seconds	20.0	5	12	31	55	21
135. Differentiates scribble from stroke	20.5	28	32	40	60	37
136. Sentence of 2 words	20.6	12	18	45	47	29
137. Pink board: completes	21.1	10	20	38	33	24
138. Names 2 objects	21.4	0	8	20	18	11
139. Points to 5 pictures	21.6	0	1	9	15	5
140. Broken doll: mends approximately	21.9	7	14	17	28	15
141. Names 3 pictures	22.1	0	0	15	18	7
142. Blue board: places 6 blocks	22.4	6	15	23	37	19
143. Builds Tower of 6 cubes	23.0	2	9	6	27	9
144. Discriminates 2 objects	23.4	0	17	21	22	15
145. Names 4 1/2 pictures	23.8	0	3	9	11	6
146. Names 3 objects	24.0	0	2	7	8	4
147. Institutes vertical and horizontal strokes	24.4	9	9	19	22	14
148. Points to 7 pictures	24.7	0	0	0	10	2
149. Names 5 pictures	25.0	0	0	2	13	3
150. Names 2nd picture	25.2	0	2	5	10	4
151. Pink board: reversed	25.4	6	5	5	19	7
152. Discriminates 3 objects	25.6	0	7	6	10	6
153. Broken doll: mends exactly	26.1	2	2	8	7	5
154. Train of cubes	26.1	2	2	2	20	5
155. Blue board: completes in 150 seconds	26.3	2	2	5	7	4
156. Pegs placed in 22 seconds	26.6	0	0	3	13	3
157. Folds paper	27.9	0	0	4	14	3
158. Understands 2 prepositions	28.2	0	0	2	0	1
159. Blue board: completes in 90 seconds	30.0	0	0	7	4	3
160. Blue board: completes in 60 seconds	30+	0	0	6	0	3
161. Builds tower of 8 cubes	30+	2	0	4	8	3
162. Concept of one	30+	0	0	0	2	1
163. Understands 3 prepositions	30+	0	0	0	0	0

Table 3-3

Percentage Passing PDS Items

Item		Normed age	15-16 months N=115	17-18 months N=111	19-21 months N=100	22-24 months N=40	All children N=366
47.	Stands up: I	12.6	91	97	92	95	94
48.	Throws ball	13.3	90	94	97	98	94
49.	Walks sideways	14.1	76	87	82	82	82
50.	Walks backward, few steps	14.6	81	91	88	93	87
51.*	Stands on right foot with help	15.9	61	69	75	93	71
52.*	Stands on left foot with help	16.1	61	71	77	90	71
57.*	Stands up: II	21.9	21	38	51	51	38
58.*	Stands on left foot	22.7	28	38	47	53	39
59.*	Jumps off floor, both feet	23.4	5	13	24	55	18
60.*	Stands on right foot alone	23.5	23	29	41	53	33
61.*	Walks on line, general description	23.9	35	36	53	73	45
65.*	Walks on tiptoe, few steps	25.7	8	8	21	35	15
68.*	Walks backward 10 feet	27.8	4	22	18	24	16
71.*	Stands up: III	30+	4	4	11	17	7
73.	Walks on tiptoe, 10 feet	30+	0	1	4	0	1
75.	Walks on line 10 feet	30+	1	0	0	0	0

*Items combined into a Physical Score that were used in analyses.

3.1.3 The Test Environment

At the conclusion of each Bayley assessment, a Child Behavior Record was completed. This record describes the test environment, as well as the social behavior of the child (and parent, if present). The results are summarized in Table 3-4. As shown, most children engaged willingly in the test and were friendly toward the examiner, with some initial wariness. A few behaviors differed by age, site, and group:

- Easy acceptance of the test was more frequent in older children (62%); frequent verbalization was observed more often in older children (30%) than in younger children (11%).
- Far more parents in Jackson, Las Vegas, and New Haven assisted children frequently (27%) than in Oklahoma City, St. Petersburg, or Salem (4%).
- More CFRP children were judged as easily distracted (23%) than control/comparison group children (14%); and 18% of CFRP parents frequently assisted or interfered in the test administration compared to 9% of control/comparison group parents.

Problems noted in the test environment were the presence/interference of other children (15%); interruptions due to the flow of people in the home (11%); noise due to television, stereo, or people (11%); heating or lighting problems (4%); and lack of availability of adequate testing surfaces (11%). However, 78 percent of the home test environments were free of any kind of problems, and the observed problems were not found more frequently for any age group or site, or in CFRP or control/comparison group families.

Table 3-4

Infant Behavior Record Summary

Category	Review of Responses
1. Social orientation (responsiveness to examiner)	17% hesitant; 53% accepting; 28% friendly; 2% inviting
2. Cooperativeness with examiner	8% refused many items; 36% refused or resisted on one or two items; 25% accepted tests willingly; 30% enjoyed and readily performed test items
3. Fearfulness	37% displayed no apparent fear; 40% some restraint during early portion of test; 14% moderate restraint during first half of test; 8% moderate restraint during much of the test
4. Goal directedness (persistence)	19% easily distracted; 54% fairly persistent; 27% very persistent
5. Communication skills	5% silent throughout testing; 50% occasional vocalization; 26% frequent vocalization/few words; 18% frequent verbalization
6. Parent behavior--test assistance	45% no assistance; 41% occasional assistance or interference; 14% frequent assistance or interference
7. Parent behavior--emotional support[a]	28% frequent positive support; 60% occasional positive support; 5% no comments to child; 25% occasional negative comments; 3% frequent negative comments
8. Judgment of test	8% fairly adequate; 36% average; 54% very good

[a]More than one category checked for some children.

3.2 Analytic Approach

To get a preliminary picture of the performance of CFRP and control/comparison children against a background of national norms, we computed normed scores for the mental development scale. (Such scores are computed so that the average score for children at every age will be 100; scores over 100 indicate that a child is developing more rapidly than average--performs above his or her age level--while scores below 100 indicate performance below age level.) Comparisons of physical scores to national norms could not be made because the physical development scale was not administered in its entirety.

Normed scores were not used in analyzing the effects of CFRP, primarily because of complications introduced by the age differences between CFRP and control/comparison groups, noted earlier. (This issue is discussed further in Section 3.3.) Rather, a Mental Score consisting simply of the raw number of mental items passed and a Physical Score consisting of the proportion of physical items passed formed the basis for all further work. (A proportion, rather than a sum, was used in the case of the physical items, because data on one or two items were missing for some children; of 365 Physical Scores, 348 are based on all ten items, 15 on nine items, and 2 on eight. Proportions largely eliminate any spurious differences in scores that would arise because of variations in numbers of items.)

Mental and Physical Scores were further scrutinized in several ways: First, their reliabilities were assessed. ("Reliability" denotes the degree to which the different items on a test give consistent results, and the degree to which the test gives consistent results when re-administered--that is, the degree to which it is free of measurement error.*) Reliabilities of both Mental and Physical scores were judged to be high.** Second, the variability of the Mental and Physical Scores was examined and found to be sufficient to suggest that further analysis to determine its sources would be fruitful. Third, the data were screened for "outliers"--children who were significantly older than the rest of the sample at the time of assessment, or those whose scores were suspiciously high or low in light of their age. (Such cases were omitted from further analysis.) Finally, relationships (correlations) between Mental and Physical Scores (Table 3-5) and between children's ages at the time of assessment and their Mental and Physical Scores (Table 3-6) were examined, within sites and within CFRP and control/comparison groups, in an effort to detect anomalous relationships that might warrant special analytic treatment. Most of the observed correlations were high, as expected; however, the results also pointed to one anomalous subsample (in St. Petersburg), which was deleted from some later analyses.

In order to compare the Physical and Mental Scores of CFRP and control/comparison children, it was necessary to adjust or control for differences between the two groups in age at testing, as well as differences in family background characteristics noted earlier. Two different techniques were used to make the necessary adjustments. One, called analysis of covariance, or ANCOVA, is rather like a handicap

---

*Reliability is estimated by an internal consistency measure, alpha. Alpha varies from 0.00 to 1.00. An alpha of 0.00 indicates a completely unreliable measure; an alpha of 1.00 indicates a measure that is perfectly reliable. Thus, an alpha of .95 indicates that 95 percent of the measured variation among the scores of children is attributable to genuine differences, and that only 5 percent of the measured variation is due to random effects or measurement error.

**Alphas for Mental Scores ranged from a low of .88 in Oklahoma City and St. Petersburg to a high of .95 in Las Vegas. Alphas were somewhat lower for Physical Score, ranging from a low of .72 in New Haven to a high of .85 in Las Vegas.

Table 3-5

Mental and Physical Score Correlations

	N		Correlation	
	CFRP	Comp.	CFRP	Comp.
Jackson	26	16	.06	.55
Las Vegas	29	21	.59	.74
New Haven	21	10	.01	.87
Oklahoma City	23	37	.63	.37
St. Petersburg	29	36	.64	.14
Salem	29	35	.34	.59

Table 3-6

Correlations of Child Age With Mental and Physical Scores

	Mental Score				Physical Score			
	N		Correlation		N		Correlation	
	CFRP	Comp.	CFRP	Comp.	CFRP	Comp.	CFRP	Comp.
Jackson	30	16	.43	.70	26	16	-.06	.79
Las Vegas	32	25	.75	.73	29	21	.64	.60
New Haven	21	9	.15	.66	21	10	.49	.65
Oklahoma City	25	37	.80	.72	23	37	.66	.38
St. Petersburg	29	31	.70	-.01*	29	36	.52	.17
Salem	27	33	.82	.83	29	35	.18	.62

*The correlation between Mental Score and age within the St. Petersburg control/comparison group is somewhat anomalous. Ordinarily, high correlations between age and Mental Scores are expected, although these may be lessened when children receive special treatment--as in the case of the Jackson CFRP group. Closer examination showed that the St. Petersburg correlation was not an error but was due to a group of young children who received apparently genuine, very high scores. Because of this anomaly, St. Petersburg was excluded from many comparisons between CFRP and control. New Haven was also excluded because the correlation within the CFRP group seemed too low to be explained by a treatment effect.

in golf or bowling: before individuals' scores are compared, an adjustment is made to compensate for known or suspected inequities in initial skill. The second, called the value-added technique, uses data from the comparison group to predict how CFRP children with given characteristics would develop in the absence of "treatment." The difference between their actual and predicted performance is the "value added" by CFRP. These analytic techniques are discussed in more detail in Appendix B. They are mentioned here to stress the point illustrated concretely later, that results are <u>not</u> always identical when different analytic approaches are used.

All analyses were conducted both <u>within</u> sites--to examine differences due to program implementation, client populations, and other site characteristics--and <u>across</u> sites--to maximize statistical power and assess the impact of the program as a whole.

3.3 CFRP's Impact on Child Development

In assessing CFRP's impact on the development of infants and toddlers, we first compared the Mental Scores of CFRP and control/comparison group children against national norms. As noted earlier, similar comparisons could not be made for Physical Scores because the physical development scale was not administered in its entirety. The CFRP and control/comparison group children combined scored slightly below national norms (97.2 compared with 100) in terms of their mental development; the results were quite similar across sites (Table 3-7). Performance with respect to norms was less satisfactory for older children at all sites. This finding parallels results from a number of other studies of infant intervention programs, including the Parent-Child Development Center (PCDC) program.

Table 3-7

Mean Normed Mental Score Performance[a]

	N	15-16	17-18	19-21	22-24	Total
		Age in Months				
Jackson						
CFRP	30	102.3	100.8	84.2	71.3	94.8
Comparison	16	105.2	98.8	94.0	n.a.	100.8
Las Vegas						
CFRP	33	104.0	94.6	91.6	87.0	96.8
Comparison	25	104.0	97.7	95.3	90.0	98.5
New Haven						
CFRP	22	118.4	98.0	87.8	73.0	93.2
Comparison	10	115.0	114.0	94.0	85.7	101.8
Oklahoma City						
CFRP	25	103.1	92.7	84.0	79.6	91.3
Comparison	38	113.6	97.0	89.6	85.8	97.0
St. Petersburg						
CFRP	28	103.3	81.3	96.8	85.3	96.0
Comparison	32	115.6	99.8	90.2	n.a.	104.4
Salem						
CFRP	28	110.2	105.2	101.7	91.7	104.3
Comparison	34	106.8	100.0	94.2	87.6	99.9

The data presented in Table 3-7 suggest that the scores of control/comparison children in most age groups and at most sites are higher than those of children in the CFRP group. This conclusion is erroneous because of the differential distribution of child ages both across and within sites. Children in the control/comparison group were generally younger than those in the CFRP group, and normed scores decrease as children get older. Also note that apparent group differences are virtually nonexistent when average Mental Scores are computed for each group by site (Table 3-8). Similar descriptive information concerning Physical Scores is presented in Table 3-9.

Mental and Physical Scores adjusted for child age and other important family background characteristics were used to assess CFRP's impact. Across-site analyses using both the ANCOVA and value-added analytic techniques provide no evidence that CFRP has had a positive impact on either the mental or physical development of children. The results are summarized in Tables 3-10 and 3-11. Three sets of data are presented in the tables. The first two columns are results from ANCOVA analyses. The net score change simply reports how many points higher (or lower)--on average--CFRP children scored on Mental Score compared with children in the control/comparison group. On Physical Score one-tenth of a point represents one test item (this is because proportions rather than number of items were used to compute this score). The change in growth rate indicates the degree to which children's scores grow more rapidly than normal as a result of participation in CFRP. The third column reports results of the value-added analytic approach; results are similar to net score changes. None of the results reported in either table were statistically or substantively significant. (Details concerning these analyses are presented in Appendix C.)

Table 3-8

Mental Score Descriptives

	N		Mean		S.D.	
	CFRP	Comp.	CFRP	Comp.	CFRP	Comp.
Jackson	30	16	121.9	122.0	5.61	7.60
Las Vegas	32	25	121.7	123.5	7.35	8.70
New Haven	21	9	124.8	125.1	5.78	5.90
Oklahoma City	25	37	122.1	123.0	5.95	5.40
St. Petersburg	29	31	123.6	121.8	7.70	2.65
Salem	27	33	126.3	123.6	7.84	2.11

Table 3-9

Physical Score Descriptives

	N		Mean		S.D.	
	CFRP	Comp.	CFRP	Comp.	CFRP	Comp.
Jackson	26	16	.32	.39	.16	.26
Las Vegas	29	21	.41	.45	.21	.26
New Haven	21	10	.60	.60	.22	.26
Oklahoma City	23	37	.57	.58	.22	.22
St. Petersburg	29	36	.49	.44	.18	.21
Salem	29	35	.60	.58	.21	.20

Table 3-10

CFRP Effects on Mental Score Across Sites

	ANCOVA		
	Net Score Change	Change in Growth Rate	Value Added
Effect	-0.29	0.01	-0.25
S.E.	0.38	0.16	0.69
p-value	>.50	.49	>.50

Table 3-11

CFRP Effects on Physical Score Across Sites

	ANCOVA		
	Net Score Change	Change in Growth Rate	Value Added
Effect	-.002	-.009	-.019
S.E.	.014	.006	.019
p-value	>.50	>.50	>.50

Finally, within-site analyses were performed on Mental and Physical Scores. Only Mental Score analyses using the predictive (value-added) approach in Salem suggest that CFRP has had an impact (Table 3-12), with a net gain of about two points on Mental Score. At eighteen months, a net average gain of two points would put the CFRP children at or above the Bayley Mental Score norm, rather than just below it. No CFRP impact on children's physical development was detected at any of the sites.

Table 3-12

Average CFRP Effects on Mental Score, Within-Site Value-Added Analyses[a]

	Jackson	Las Vegas	Oklahoma City	Salem
Effect	-1.78	-0.71	-1.86	2.18
S.E.	0.94	0.87	0.63	0.87
df[b]	28	28	22	26
p-value	>.50	>.50	>.50	.04

[a]New Haven and St. Petersburg excluded: see note attached to Table 3-6 for an explanation.

[b]Significance tests controlled for four simultaneous one-tailed tests.

3.4 CFRP Participation and Child Development

Although there are no significant differences, on average, between the Bayley developmental scores of CFRP and control/comparison group children, we cannot conclude that CFRP has no effect on children's development. CFRP is not a uniform, standardized program; services are tailored to the needs and preferences of families (that is, they are "individualized"), and there is considerable variation among sites in program structure and activities.* Thus different

*See Phase III Program Study Report, August 1980.

families and children "participate" in CFRP in different ways. Moreover, as shown in Chapter 2, families differ widely in their frequencies of participation in program activities. It is quite possible that these differences in type and amount of participation are related to the developmental benefits children derive from the program. This overall comparison may be concealing program effects that occur only when the child participates in certain ways, or participates often enough or long enough.

Numerous analyses were performed in order to explore these possibilities. Details are reported in Appendix D; major findings are summarized here. On balance, results suggest that developmental benefits, as measured by the Bayley mental and physical scales, may in fact relate to the amount and type of the family's participation in the program, although this conclusion is clouded by differences in participation rates from site to site.

3.4.1 Effects of Amount and Rate of Participation

The simplest view of the relationship between participation and developmental benefits is that "more is better"--more participation will lead to greater benefits. (There are reasons why this view might not hold for CFRP; we shall examine these reasons later.) There are three types of CFRP activities to which the maxim "more is better" might apply: parent education center sessions, infant-toddler center sessions, and home visits. As shown in Chapter 2, all families receive home visits, but with varying frequency. Many do not attend center sessions at all, and, among those who do, there is substantial variation in the frequency of their attendance. Consequently, effects of monthly rates of participation in parent education sessions, infant-toddler sessions, and home visits were examined separately.

Summary statistics for these three measures are presented in Table 3-13.*

The participation measures were examined in relationship to five different measures of growth: Bayley scale mental growth indexes calculated (for the entire sample) by the value-added and ANCOVA techniques, physical growth indexes calculated for the entire sample by both techniques, and a mental growth index calculated by the value-added technique for the Salem sample. (Recall that value-added scores in Salem were the only mental growth indexes at any site to show a positive overall effect of CFRP enrollment.)

Table 3-13

Summary Statistics, Participation Measures[a]
(per month)

	Minimum	Maximum	Median	Mean	S.D.	N
Home Visit Rate	0.0	3.9	1.1	1.4	0.8	143
Infant-Toddler Session Rate	0.1	2.2	1.0	1.1	0.5	61
Parent Education Session Rate	0.2	1.7	1.0	0.8	0.4	66

[a]For each family, participation measures cover the period preceding testing of the child.

*The Bayley scales had to be administered across a period of six months in order to ensure that all children would be at least 15 months old when tested. Consequently, there were appreciable differences in amounts of time that different families had been in CFRP prior to testing. Participation data used in all analyses in this chapter reflect these differences; participation rates for each family are based only on the period preceding testing. In addition, the analytic sample included all available families, not just those whose participation rates for various activities exceeded once per quarter. Consequently monthly participation rates shown in Table 3-13 are not directly translatable into the quarterly rates presented in Chapter 2, which cover participation over the same time period, regardless of when the child was tested.

The two physical measures were so highly correlated (.87) as to be almost indistinguishable. The two types of mental measure were also highly correlated (.68), but were sufficiently distinct that they might prove sensitive to different participation measures.

Relationships between the three participation indexes and the five developmental indexes are shown in Table 3-14. There are enough significant and near-significant correlations in the table to lend credence to the general view that program participation affects both mental and physical development of children. Specifically, although the pattern of findings is quite complex, it suggests that attendance at infant-toddler center sessions is positively associated with mental growth as measured by the ANCOVA method and with physical growth measured by both methods.

However, caution must be used in interpreting the association between participation and the developmental measures. Recall from Chapter 2 that there are substantial differences in participation from site to site. Because of these differences it is possible that what appears to be a relationship between <u>participation</u> and developmental benefits is instead a relationship between <u>site</u> and developmental benefits. The only definitive way to disentangle these rival interpretations would be to determine whether there is a relationship between developmental benefits and participation <u>within</u> sites. Unfortunately, sample sizes within sites do not offer enough statistical power to perform such analyses with any confidence. Thus both interpretations remain equally consistent with the data; this fact must be kept in mind in reading the results of more detailed analyses of participation for the sample as a whole, which are reported below.

Table 3-14

Correlations Between Participation Measures and Mental and Physical Growth Indexes

	Home Visit Rate	Infant-Toddler Session Rate	Parent Education Session Rate	Center Partici-pation/Non-participation
Mental Score Value-Added	-.02	.10	-.05	.06
N	91	41	43	94
p-value[a]	>.50	>.25	>.50	>.25
ANCOVA Mental Score Residual	.06	.39	.05	-.00
N	97	40	44	100
p-value[a]	>.25	.006	>.35	>.50
Within-Salem Mental Score Value-Added	.02	.19	.20	.25
N	27	18	12	27
p-value[a]	>.45	>.23	>.25	>.10
Physical Score Value-Added	.11	.32	-.07	.20
N	105	46	47	108
p-value[a]	>.10	.015	>.50	.017
ANCOVA Physical Score Residual	.17	.27	-.16	.13
N	98	39	47	101
p-value[a]	.044	.049	>.50	.09

[a]Significance levels are given for individual, one-tailed tests. As is well known, individual significance tests can be misleading when many tests are performed at once. (Five out of 100 will achieve conventional "significance" levels by sheer chance.) Across five simultaneous tests, a univariate significance level of .02 is actually significant at .10. Across twenty-five simultaneous tests, a univariate significance of about .004 is required for overall significance at .10.

3.4.2 Interactions Among Types of Participation

Closer examination of the simple hypothesis that "more (participation) is better" suggests that it may be too simple for CFRP. The various program elements are intended to work in concert with one another, not in isolation. Thus, for example, it might be the case that the combined impact of frequent home visits and frequent center attendance is far greater than the sum of the two effects taken separately. Statistically, this synergistic combination of effects is an "interaction."

One way to test for the presence of interactions is to form the product of the two participation measures in question and examine the correlations of these products with the various growth measures. Table 3-15 shows relationships between the growth indexes and the combinations (interactions) of home visit rates with the two center participation indexes. Thus, for example, the upper left-hand cell of the table shows the combined effect of home visit rates and infant-toddler session rates on the Bayley mental growth index, calculated by the value-added technique. Several correlations are significant or near-significant, despite the fact that the need to combine variables for this analysis resulted in a reduction of sample size, hence of significance level, for each correlation. Thus, the table points to the possible importance of the combined effects of home visits and infant-toddler sessions, as well as the combined effects of home visits with some center participation on the part of parents. Subject to the qualification about site differences raised above, participation in center sessions, particularly infant-toddler sessions, seems to promote mental and physical development--and home visits appear to interact with center participation to give an added "boost" to Bayley scores.

Table 3-15

Correlations Between the Combined Effects of Home and Center Participation and Mental and Physical Growth Indexes

Combined Effect (Interaction) of Home Visit Rate and:

	Infant-Toddler Session Rate	Parent Education Session Rate	Center Partici-pation/Non-participation
Mental Score Value-Added	-.02	-.05	.03
N	40	41	91
p-value[a]	>.50	>.50	>.35
ANCOVA Mental Score Residual	.41	.05	.00
N	39	42	97
p-value[a]	.005	>.35	.50
Within-Salem Mental Score Value-Added	.18	.10	.25
N	18	12	27
p-value[a]	>.20	>.35	.104
Physical Score Value-Added	.13	-.17	.12
N	45	45	105
p-value[a]	>.15	>.50	.106
ANCOVA Physical Score Residual	.24	.01	.15
N	38	45	98
p-value[a]	.074	>.45	.076

[a]Significance levels are given for individual, one-tailed tests.

3.4.3 Effects of Family Needs and Strengths

Still further consideration of the nature of CFRP suggests that the meaning of "participation" is conditioned on the needs and strengths of families. Because families participate voluntarily, they are likely to seek out the program in times of particular need, and they are likely to ask for help in areas where their needs are greatest. Conversely, they are least likely to seek help in areas of their greatest strength. Thus, a casual (and misleading) analysis might show a negative relationship between program participation and the family's level of functioning. Clearly, to assess the effects of participation, it is necessary to consider families' needs and strengths as well.

The amount of benefit that a child experiences on some outcome measure may also depend on the needs and/or strengths of the child's family. It might be the case that children from the neediest families profit the most from CFRP. On the other hand, a certain amount of family support (strength) may be necessary for a child to benefit from the program. Both hypotheses are plausible; they imply opposite relationships between family needs/strengths and the degree to which CFRP benefits children, but they also imply that family needs and strengths cannot be ignored in assessing CFRP's impact on the child.

To begin to approach the difficult problem of measuring the needs and strengths of families, we constructed a set of nine measures, based on parent interview data and staff reports. The nine measures are listed and defined in Table 3-16; basic descriptive statistics are also shown. The first two measures--HASSLED and NEEDS--are fairly straightforward, though global, measures of need. The next three measures--CFRP SUPPORT, RESOURCES and ENTHUSIASM--are equally straightforward measures of strengths. CD EMPHASIS is a less

Table 3-16

Needs and Strengths Measures

Name and Description	Summary Statistics
HASSLED--the extent to which one feels "hassled" generally, by people in various roles (doctor, neighbors, family, friends, and relatives). From fall 1978 parent interviews. (A higher score is more hassled.)	min = 0.00 max = 1.00 mean = 0.52 median = 0.50 SD = 0.34 N = 143
NEEDS--an index of extent of family needs across many general areas. From fall 1978 and spring 1979 staff reports. (A higher score indicates more needs.)	min = 0.39 max = 2.86 mean = 1.55 median = 1.60 SD = 0.50 N = 126
CFRP SUPPORT--the degree of support for CFRP involvement from family and friends. From fall 1978 and spring 1979 parent interviews. (A higher score indicates greater support.)	min = 1.11 max = 3.46 mean = 2.31 median = 2.32 SD = 0.74 N = 126
RESOURCES--an assessment of the personal or social resources available to the parents, such as frequency of social contacts outside of CFRP, ties with extended family, and awareness of social services available in the community. From spring 1979 staff reports. (A higher score indicates greater resources.)	min = 0.25 max = 1.67 mean = 0.93 median = 1.00 SD = 0.33 N = 126
ENTHUSIASM--the extent of interest in or enthusiasm for CFRP activities, across a number of specific items. From spring 1979 staff reports. (A higher score shows greater enthusiasm.)	min = 1.01 max = 3.53 mean = 2.29 median = 2.30 SD = 0.68 N = 130
SPOUSE HELPS--the extent of help from one's spouse (or live-in partner) in routine child care tasks. From fall 1978 parent interviews. (A higher score means more help.)	min = 0.13 max = 1.00 mean = 0.50 median = 0.50 SD = 0.24 N = 68[a]
RELATIVE HELPS--like SPOUSE HELPS, but for other relatives. From fall 1978 parent interviews. (A higher score means more help.)	min = 0.13 max = 1.00 mean = 0.58 median = 0.62 SD = 0.29 N = 89[a]
OTHERS HELP--like the two measures above, but for other people helping; the natural logarithm of this variable was used. From fall 1978 parent interviews. (A higher score means more help.)	min = -2.08 max = -0.13 mean = -1.58 median = -1.38 SD = 0.56 N = 68[a]
CD (CHILD DEVELOPMENT) EMPHASIS--a three-valued assessment of the extent to which child development was emphasized during the first six to eight months of CFRP. From spring 1979 staff reports. (Codes: 0 - not at all; 1 - somewhat; 2 - fairly strongly.)	0 - 58 families 2 - 22 families 1 - 55 families N = 135

[a]The relatively high incidence of missing data on the HELPS constructs seems to reflect either the unavailability of or the failure to use others in helping with routine child care chores, or some combination.

straightforward measure of need; staff who placed special emphasis on child development during the first months of the program appeared to be responding to perceived special needs on the part of children. The remaining measures--the HELP series--may reflect needs, strengths, or both.

We then used these nine measures as "covariates" in a series of exploratory multiple regression analyses. Roughly, this statistical approach affords a means of holding needs and strengths constant in order to isolate the effects of participation on mental and physical growth. On the whole, taking account of needs and strengths did not change the pattern of associations between participation and growth indexes that resulted from the simple analyses reported earlier (Table 3-14). Regression does, however, add something to simpler correlational analyses, in that it allows us to estimate the magnitudes of the increments in growth resulting from different levels of participation: An increment of one infant-toddler session per month produced an increment of about 3 points in Bayley Mental Score, in the ANCOVA model (p=.06). Similarly, one additional infant-toddler session per month produced an increment of one item on the physical scale in either model (p=.05 for value-added; p=.10 for ANCOVA).

Additionally, more complex regression analyses explored the effects of various combinations of participation variables, in tandem with the measures of needs and strengths. Among the salient results of these explorations were the following: (1) Greater RESOURCES (defined in Table 3-16) were associated with higher mental growth indexes. (2) When RESOURCES were held constant, the combination (interaction) of infant-toddler sessions and home visits was still a potent one: one additional home visit per month seems to add about 1 1/2 points on the Bayley mental scale to the increment in growth produced by infant-toddler session attendance alone, and an additional infant-toddler session per month seems to add about 1 1/2 points to the increment produced by home visits alone. (3) Children whose mothers had greater RESOURCES profited more from infant-toddler sessions than those whose mothers had few personal resources.

3.5 CFRP and Child Development: Summary and Conclusions

No average difference in mental and physical development, as measured by the Bayley scales, was found between CFRP and control/comparison children after the former had participated in the program for 12 to 18 months. Excepting the suggestion of a possible effect on Mental Scores at one site, CFRP has not (or not yet) produced more rapid or sustained development than could be expected in the absence of CFRP enrollment. However, this result does not necessarily mean that CFRP will never show effects, nor does it mean that a program like CFRP cannot enhance development in very young children.

In an effort to explore further the potential of CFRP to enhance the development of infants and toddlers, we investigated the effects of program participation on indexes of mental and physical growth. The logic of our investigation was that the program had its best chance to work among families who participated actively--and we knew that many did not. The investigation bore fruit. Hints of effects of participation were scattered across five outcome measures, but all had a common thread: Participation in infant-toddler sessions seems to be a key contributor to the impact of CFRP on Bayley scores. Whether alone, in interaction with home visits, or in interaction with the personal resources available to CFRP mothers, the rate of participation in infant-toddler sessions is positively related to both mental and physical gains.

It is important not to overstate these findings. We have summarized here only those results that point to potential effects of CFRP participation. Other analyses, giving no hint of CFRP effects, have been ignored in this summary. In addition, as already pointed out, the possible effects of participation are confounded with site differences. Nonetheless, the pattern of results is suggestive.

Setting aside these necessary caveats, the results appear to have several implications for the operation of CFRP. The sobering and overriding conclusion is that, while there are indications that the program has the <u>potential</u> to enhance the development of children under age two, CFRP is a long way from realizing that potential. Typical participation rates are too low for the program to have much effect. (In the current evaluation sample, roughly two-thirds of the CFRP families attended infant/toddler sessions less than once per quarter during their first 18 months of enrollment. Among those who did participate, the median rate of attendance was once per month.) To increase the effectiveness of the program, ways must be found to increase attendance at infant-toddler sessions. The effectiveness of such participation can be "levered" by frequent and regular home visits. Finally, consonant with CFRP's philosophy of working through the family, increasing the personal and social resources available to the mother--e.g., increasing her awareness of community services, or helping her establish and maintain ties with neighbors and her extended family--can further "lever" the developmental effects of home visits and infant-toddler sessions. All of this is more easily said than done, but appears to be necessary if the program is to achieve its goals for children.

Chapter 4

EXECUTIVE SUMMARY

CFRP was designed "to assist parents to promote the total development of children." A summary of findings concerning the impact of CFRP's infant-toddler component on the development of children between the ages of 15 and 24 months is presented below.

Question 1: Is CFRP effective in promoting the development of infants and toddlers?

NO. When tested after a year to a year and a half of program participation, CFRP children were not found to differ significantly from control/comparison children on mental and physical development scores of the Bayley Scales of Infant Development (BSID). There is an indication of positive CFRP impact on mental development at only one site.

One possible explanation is that family participation in infant-toddler activities, particularly center sessions, is "less than optimal." There also is some question as to whether home visits occur with sufficient frequency to carry out an effective infant-toddler program in the home. Frequency is dictated to a large extent by high family worker caseloads, exceeding 20 families at some sites.

Question 2: Is family participation in infant-toddler program activities related in any way to child development?

YES, apparently. Rate of participation in infant-toddler center sessions is positively associated with both mental and physical development scores of the BSID. The evidence also suggests that the benefits children derive from these sessions may be related to rate of participation in home visits and to the availability of personal and social resources to the mother or family outside of CFRP. That is, the positive effect of these sessions tends to be greater for families with higher participation in home visits and greater availability of resources.

It is important to point out that just about half of the CFRP families had never attended an infant-toddler center session as of the date of Bayley testing. Among those who had attended one or more sessions, the median rate was one session per month (during the period in which sessions were attended). The finding reported above suggests that more active participation in program activities could result in demonstrable differences between the CFRP and control/comparison groups on the mental and physical development of children, favoring the CFRP group.

There are several other issues which should be considered in connection with CFRP's effect--or lack of effect--on child development scores as measured by the Bayley scales. First of all, there is some question as to the appropriateness of this instrument for measuring CFRP impact. Further, it is possible that CFRP impact on child development is simply too indirect to be detected after 18 months of program involvement. Parents may have to change first, before their children are likely to be affected. If positive findings (of CFRP impact) from a pilot study involving observations of parent-child interaction are replicated--that is, if CFRP participation is effecting genuine positive changes in parent-child interaction--there is reason to believe that the eventual result will be (measurably) enhanced development of CFRP children.

If CFRP is not significantly affecting child development, there are several possible (and plausible) explanations. Amont them are: typical CFRP approaches to parent education (relying on didactic methods more than on observation of modeled behavior and opportunities to practice newly acquired parenting techniques); the typical focus of home visits (on family needs rather than parenting and child development concerns, perhaps partly associated with a lack of child development background on the part of family workers); the low frequency of home visits and the low levels of family participation in center sessions. Changes in educational approaches and shifts in home visit focus would presumably do much to strengthen CFRP as a child development program; however, the second finding reported above suggests that increases in the frequency of contact between program and family might well do even more.

Appendix A

BAYLEY DATA QUALITY

It is customary in large-scale data collection efforts, especially those involving assessments of child development, to gather information concerning the quality of data obtained by on-site research staff. Tester performance in test administration and scoring was monitored on an ongoing basis during the fall/winter 1979-80 Bayley data collection period. In this appendix the monitoring process is described and the quality of the Bayley data reported.

The monitoring process started during tester training, when a trainer accompanied trainees on several practice visits, designed to evaluate trainee performance. Additional training and assistance was provided to trainees on the basis of these evaluations to ensure uniform test administration. This process was repeated on site as part of the data collection start-up process. No testers were permitted to proceed with data collection tasks until their performance was deemed to be satisfactory by monitoring staff.

Once the data collection effort had gotten under way, on-site research staff were responsible for ongoing monitoring of data quality. Staff periodically accompanied each other on testing sessions. One person was responsible for both test administration and scoring of items, while the other person independently scored the child's responses and observed tester performance. At the conclusion of the session, staff compared scores in the two test booklets and reviewed any problems that occurred during the course of the test administration. Scores were not changed, however, in this review process. Completed tests and monitor booklets were reviewed a second time by a data collection supervisor in Cambridge. Discrepancies in scores were tabulated in order to assess data quality. Periodically, on-site data collection staff received feedback concerning their performance based on this review.

As is noted in Table A-1, a total of 81 Bayley testing sessions were monitored in this fashion during the fall/winter 1979-80 data collection period, an average of 22 percent of all Bayley assessments. The number of monitoring visits ranged from a low of 10 (14%) in Salem to a high of 20 (31%) in Las Vegas.

Table A-1

Monitored Tests

	N	Percent of Total Tests
Jackson, MI	11	22
Las Vegas, NV	20	31
New Haven, CT	12	30
Oklahoma City, OK	16	23
St. Petersburg, FL	12	17
Salem, OR	10	14
Total	81	22 (site average)

The Bayley consists of two scales--the mental development scale (MDS) and the physical development scale (PDS). The MDS contains a total of 69 items for children in the 15-month or older age range. The PDS is considerably shorter, with a maximum of 16 items (excluding those requiring the step stool and walking board).* The total number of items administered depends on the child's performance. An average of 32 MDS items were administered to children in the CFRP evaluation; for the PDS the mean number of items was 13.

*See Chapter 1 for an explanation on why certain items were excluded from the PDS.

In order to assess Bayley data quality, we computed the mean number of scoring discrepancies that occurred on each Bayley test, as recorded in monitor booklets. An average of 2.2 discrepancies were found on the two Bayley scales combined, an overall error rate of less than five percent. Four sites had fewer discrepancies per test, as noted in Table A-2 (Las Vegas, Jackson, Oklahoma City, and Salem); the Bayley data obtained in New Haven and St. Petersburg were of considerably lower quality, with per-test discrepancies averaging 5.4 and 3.2 respectively. One of the New Haven testers was terminated on the basis of these data for the January-through-March data collection period. Scoring discrepancies were more common on the MDS than on the other scale, due largely to differences in the number of items in each scale.

Table A-2

Mean Number of Discrepancies per Test

	MDS	PDS	Total
Mean # of Items	32	13	45
Jackson, MI	.9	.3	1.2
Las Vegas, NV	.7	.3	1.0
New Haven, CT	4.5	.9	5.4
Oklahoma City, OK	.7	.8	1.5
St. Petersburg, FL	2.8	.4	3.2
Salem, OR	1.2	.3	1.5
Site Average	1.7	.5	2.2

To examine the sources of scoring discrepancy more closely, we singled out portions of the test for which discrepancy rates were particularly high. Items on the Bayley are administered in the context of different situations. Each situation contains items that are similar. In the "Cube Behavior" situation (part of the MDS), for example, the child is first asked to unwrap a cube. If the item is passed, the child is asked to build a tower of 2, 3, 6, and 8 cubes. Other related items concern building a train of cubes and using a single cube to demonstrate understanding of the concept of one. Items are similarly grouped in situations on the PDS. Discrepancy rates varied substantially across situations.

The MDS contains 19 situations. Table A-3 identifies nine situations on which scoring reliability was relatively low. The figures reported in the last column were derived by adding the total number of discrepancies that occurred across sites on all items in each of the situations and dividing this total by the number of items contained in the situation. The figures thus represent average total discrepancies per item across the six sites.

Table A-3

Discrepancies by Situations

MDS

Situation and Item Descriptions	# of Items in Situation	Average Total Discrepancies per Item
Jointed doll (follows directions and points to parts of body)	2	5.0
Vocalizations and words* (jabbers, says words, uses words to make wants known)	4	4.3
Misc. items (pushes car, dangles ring, uses gestures to make wants known,* attains toy with stick, and finds 2 objects)	5	3.2
Verbal comprehension (imitates word,* shows shoes, other clothing or toy)	2	3.0
Mends broken doll	3	3.0
Names objects (ball, watch, pencil, scissors, cup)	3	2.7
Names and points to pictures	6	2.5
Peg board (places pegs in board)	5	2.2
Pink board (places shapes in puzzle)	3	2.0

The situations reported in Table A-3 account for 81 percent of all discrepancies. On all other situations, the total number of discrepancies per item was less than 2. Three of the nine situations in Table A-3 (indicated by asterisks) involve judgments of the child's behavior on the part of the tester, rather than direct responses of the child to directions or questions. Unreliability of judgments of relatively unstructured behavior thus appears to be a contributing factor in the overall discrepancy rate.

On the PDS, items combine into six situations. In five of the situations, the total discrepancy rate per item exceeded 2.0, as noted in Table A-4.

Table A-4

Discrepancies by Situations

PDS

Situation and Item Descriptions	# of Items in Situation	Average Total Discrepancies per Item
Walking skill--pull toy (sideways, backward)	2	3.5
Walks on line (forward, backward, on tiptoe)	5	2.6
Balance (stands on one foot with help and alone)	4	2.5
Stands up from floor alone	3	2.3
Jumps from floor	1	2.0

Summary

The overall discrepancy rate per test was low, even on situations with generally lower scoring reliability. Scoring discrepancies on the jointed doll situation (the single least reliable situation), for example, occurred on only one out of eight tests. The factors examined lead to the conclusion that the Bayley data were of high quality. Possible exceptions are found at two sites--New Haven and St. Petersburg--which had a higher than average per-test discrepancy rate.

Appendix B

TESTS OF IMPACTS, MODELS, AND ANALYTIC STRATEGIES

Any statistical test of a program's impact involves an implicit or assumed model of how that program influences the outcome being measured. Further, when the outcome is a developmental phenomenon (that is, some kind of growth or change which is expected to occur "naturally" and systematically over time), the test employed makes very strong assumptions about the growth pattern and/or the way in which a program affects growth. In many evaluation applications, implicit models of impact and/or growth are not scrutinized carefully, resulting in tests based on questionable assumptions. At best, such tests are inappropriate, and the significance levels reported are wrong. At worst, bias in the tests may provide misleading results: an effect may be inferred that in fact does not exist, or an important effect may not be detected even when the research design appears to have sufficient power to detect an effect of practical significance.

Section B.1 of this appendix examines the models and assumptions implicit in a relatively common statistical testing procedure--analysis of covariance, or regression adjusted tests. Another, more recently developed approach, "value-added" modeling (see Bryk and Weisberg, 1976a and 1976b, and Bryk, Strenio, and Weisberg, 1980), is examined in Section B.2. The rationale for choosing the tests reported elsewhere in the volume is laid out in some detail in Section B.3. While a number of subtle and complex problems are reviewed here, little mathematical detail is included; references to more thorough, more rigorous discussions are given.

B.1 Traditional Tests

To understand the model and assumptions underlying traditional statistical tests of impact, consider a basic, one-way ANCOVA (lower case characters represent population parameters; upper case characters are either variables or numbers; subscripts perform their usual indexing function):

$$[1] \qquad Y_{ij} = m + g_j + b_1X_{i1} + b_2X_{i2} + \ldots + b_KX_{iK} + e_{ij}$$

Simply stated, the score of individual i in group j (Y_{ij}) consists of a grand mean (m) plus the group's deviation from the grand mean (g_j; j = 1, ..., J groups) plus a multiple regression contribution reflecting the effects of a set of covariates such as the individual's age, sex, etc. (the terms b_kX_{ik}; k = 1, ..., K covariates), plus an "error" term (e_{ij}). A typical null hypothesis is that the treatment group means ($\bar{Y}_j$) are equal.

The formal assumptions of the test of this typical null hypothesis are usually stated in terms of the distribution of the "error" and its relationship to other variables in the equation: e_{ij} is assumed to be distributed normally, independently of each other e_{ij} with zero mean and unknown variance; that variance is assumed to be constant across all values of the other variables in the equation (note that this subsumes the assumption of equal "error" variance within treatment groups); e_{ij} is independent of (hence, has zero covariance with) all the variables in the equation. When these conditions are met, the statistical test of the stated null hypothesis is justified and conclusions drawn from the tests are "reasonable."[1]

Most researchers are familiar with basic design strategies and diagnostic techniques for protection against blatant violations of these assumptions, given that the model is reasonable. If possible, samples are drawn randomly, and assignments to groups are made randomly; outliers are screened,[2] scatterplots are examined; and variables are transformed as appropriate. When necessary, to accommodate known violations of these assumptions--such as heteroscedasticity (conditional "error" variance itself varying according to the values taken on by other variables in the model) or serial correlation among the "errors"--estimation techniques other than ordinary least squares are employed. Such techniques include weighted and/or generalized least squares (Theil, 1971) as well as robust and resistant estimation (e.g., biweighted least squares; Mosteller and Tukey, 1977). No practical "tricks-of-the-trade," however, can overcome substantive insensitivity to the model employed, particularly in its assumptions about the process whereby the outcomes are generated.

Consider an all-too-familiar instance: A single program's effect is to be assessed via contrasts between scores in the treatment group (the program's clientele) and an available comparison group. In many such applications, a researcher or evaluator must use "naturally occurring" groups, for whatever reasons. Care may be taken to strive for a comparison sample that represents a population very much like that sampled for the program's clientele. Initially, then, a contrast in means may be thought to be sufficient for testing the program's impact. The formal model is:

[2] $$Y_{ij} = m + g_j + e_{ij},\ j = 1,\ 2 \text{ groups.}$$

The assumption that e_{ij} is distributed independently of g_j is crucial: if some other variable (e.g., race, SES, sex, education, cognitive skills, age) is related to the outcome measure Y and is differentially distributed in the program and comparison group, this assumption is invalidated and the test, clearly, is of little use.[3] (It is important to note, here, that "error" appears in quotes because the term e_{ij} contains not only random, unintelligible disturbances or deviations from expected values of Y_{ij} given j, but also all variables systematically related to Y_{ij} that are not included in the model underlying the test. This type of modeling error, the failure to include important variables that are correlated with the outcome and other independent variables in the model, is known as "specification error.")

Consider a hypothetical case. Imagine that the program in question concerns assertiveness training, and is implemented within a multinational corporation for a random sample of its mid-level executives in a large subsidiary. Because these staff interact frequently with one another, "treatment diffusion" could lessen the ability to detect an effect on assertiveness if a control group were drawn from the same population. For comparison purposes, therefore, a random sample is drawn from the mid-level executives of a second large subsidiary, engaged in a business very much like that of the first but located in another state; the executives in question have little interaction with each other across subsidiary boundaries. After completing the training program, some kind of assertiveness measure is taken on all executives in both groups.

Further suppose that the first subsidiary is a relatively new company, however, and its mid-level executives are relatively young, with an equal male-female balance. The second company (providing the comparison sample) is older, with older mid-level executives (e.g., characterized by slower but not otherwise necessarily different career growth patterns), most of whom are male. Finally, suppose further that age is positively associated with assertiveness in both samples, and that men are more assertive, typically, than women. These conditions invalidate a simple t-test for the effect of the training program: e_{ij} includes two variables (sex and age) that are associated with both assertiveness and group "assignment"; the estimated average level of assertiveness in the comparison group is not a "fair" (in this case, unbiased) estimate of the treatment group's average assertiveness level in the absence of the training program.

The situation described--both in general terms and in the specific example provided--illustrates the primary problem of the "nonequivalent control group" quasi-experimental design (Campbell and Stanley, 1963). Some kind of adjustment to the estimated treatment effect--the contrast in group means--is required to account for the bias due to initial nonequivalence of the two groups. The most common resolution of this problem is to employ an analysis of covariance rather than a simple analysis of variance (the t-test for a difference in group means is nothing more than an analysis of variance with two groups).[4] The covariates implicated in violating the assumption that the ANOVA "errors" are independent of group membership are entered into a multiple regression, "controlling" or "adjusting" for the covariates' influence while estimating the (now adjusted) contrast in group means. The formal model is that given in [1].

Next consider the substantive meaning of this expanded (ANCOVA) model. The regression coefficients are given as (that is, presumed to be) identical within each treatment condition. In the common evaluation application, with one treatment and one comparison group, the regression surfaces (or lines, if there is but one covariate) are presumed to be parallel. In other words, the net average program effect is assumed to be constant across all values of the covariate; the regression slopes--rates of change in the outcome measure per unit change in the corresponding covariate--are presumed to be unaffected by the program intervention. Frequently, this ANCOVA assumption is violated: if an expected program effect is the alteration of some relationship between an outcome and one or more covariates, the "homogeneity of regression" assumption (that of parallel regression surfaces) is untenable. This leads to a subtle but devastating violation of the assumption of independence of "errors" and covariates; the estimated treatment effect is biased--again--and conclusions drawn from the test of "adjusted" mean contrasts provided by the ANCOVA estimation are apt to be very misleading.[5]

"Heterogeneity of regression" poses a number of important problems for estimating and testing treatment effects. An effect can be defined conditionally as the difference between regression surfaces at any given set of covariate values. Presumably, such effects can be averaged over values of the covariates in which anyone is interested. If the regression surfaces intersect within this range, however, this averaging will include effects in both directions tending to understate the magnitude of effects in either direction. Furthermore, if the covariates are distributed differentially in the two groups, any such averaging should be weighted to reflect the practical importance of this differential distribution.

A different aspect of the "heterogeneity of regression" problem concerns the intended effect of the intervention. It may be that a program's goal is to alter the regression surface, creating "heterogeneity of regression" if the program is successful. Consider the assertiveness program illustration mentioned above. One objective of such an intervention could be to increase assertiveness most for younger executives, attenuating the existing relationship between assertiveness and age (presumably enabling managers to reach intended productivity levels earlier in their careers). Under these circumstances, the most appropriate test of the program's effect might be the contrast in regression slopes rather than any average of differences in the regression surfaces.[6]

When the research design is of the "nonequivalent control group" variety, there are other serious problems for demonstrating a treatment effect. While not unique to situations with "heterogeneity of regression," the problems of logical inference are complicated when such heterogeneity is present. Of particular concern here is the attribution of cause for any statistical effect found. Are the differences found (either between regression surfaces or regression slopes) due to the intervention or to other, perhaps unmeasured variables?[7]

While a number of (often overlooked) problems of ANCOVA models have been reviewed, one more deserves specific mention. This concerns developmental outcomes, where age is or is expected to be an important covariate (e.g., any developmental assessment of children). Cross-sectional modeling of any developmental phenomenon poses difficult problems for analysis and hypothesis testing. If age at testing is unknown and differentially distributed across groups, the estimate of treatment effects will be biased, the direction of bias depending upon the differential age distribution. Recording age at testing and using it as a covariate will not necessarily be the appropriate adjustment, however. The relationship of age with test scores may not be linear, and some transformation of the test scores may be required to linearize the relationship with age. Many developmental tests also exhibit some kind of variance heteroscedasticity with age; the "fan spread" phenomenon is a well-known example, where the variability of tests scores increases with the age of the children being tested.

Also, many developmental tests exhibit so-called "floor" and/or "ceiling" effects (lower and/or upper limits to the possible test scores, effective for anyone being tested at some point on the age continuum). Such effects require a nonlinear <u>and</u> heteroscedastic modeling of tests scores by age, although with restricted age samples a linear model may suffice <u>if</u> extrapolation beyond the sampled age range is not required. (For a more technical discussion of problems in modeling developmental phenomena, see Bryk and Weisberg 1976b). Finally, developmental interventions, typically, are not expected to yield constant effects at every age; the expected effect is greater with increasing age. That is, the treatment effect expected is an accelerated growth rate. If a simple increase in the regression slope (test score on age) is expected, the ANCOVA effect estimator is not g_j--a group level of scoring effect--but another regression parameter for a group-by-age interaction--a group rate of growth effect.

In summary, traditional ANCOVA testing techniques employed in "nonequivalent control group" designs (especially) employ very strong assumptions about the relationships between "errors" and other variables in the model, particularly the variable indexing the treatment/control group distinction(s). Unless these assumptions make <u>substantive</u> sense relative to the outcome(s) intended and measured, and to the covariates included in ar excluded from the analytic model, ANCOVA tests are very likely to be biased, inappropriate, and misleading as far as policy implications drawn from the ANCOVA estimates and tests are concerned. Even with a true experimental design, the ANCOVA model must make sense, substantively, or run the risk of providing misleading, irrelevant, and/or inappropriate information.

B.2 Value-Added Models

An important characteristic of most regression adjustment strategies (of which ANCOVA is the prototypical model) is that they attempt to "correct" estimates of treatment effects that are problematic (for whatever reason). An oversimplified but useful characterization of the value-added approach is that it attempts to model an "expected" outcome in the absence of intervention or treatment, and estimates the treatment effect from the differences (observed less expected-in-the-absence-of-treatment) found in the treatment group. Unlike other regression adjustment models, a value-added model will not contain an "effect" parameter that is estimated and tested routinely in a regression or ANCOVA fitting. Under certain conditions, a value-added model may become a model that is, for all purposes, an ANCOVA model.

Value-added models have found frequent application in estimating program impacts on developmental phenomena (Bryk and Weisberg, 1976a; Bryk, Strenio, and Weisberg, 1980). In its simplest form, one developmental model is:

$$[3] \qquad Y_i = a_i + b_i A_i + e_i \quad (A_i \text{ is age at testing})$$

where the parameters of interest are no longer assumed to be constant across individuals. Both the intercept (a_i) and the regression slope (b_i) are assumed to be <u>random</u>, sampled from distributions with means (m_a, m_b) and variances (S_a^2, S_b^2) respectively. If a_i and b_i are constant across individuals (that is, $S_a^2 = S_b^2 = 0$), then this becomes a familiar, simpler regression:

$$Y_i = a + bA_i + e_i$$

Similarly, if interest is centered around the mean intercept and mean slope, [3] can be rewritten in like form:

$$[4] \qquad Y_i = m_a + m_b A_i + e_i^*$$

where $e_i^* = e_i + (a_i - m_a) + (b_i - m_b)A_i$.

The value-added parameter is introduced simply as another term (v_i),

$$Y_i = a_i + b_i A_i + v_i + e_i$$

where v_i is assumed to be distributed with mean and variance (m_v, S_v^2). Often, the parameter of interest is m_v, the average value added as a result of the intervention.

Frequently, it is more realistic to assume that growth rates are related to important background characteristics; the developmental and value added models become more complicated:

$$[5] \qquad Y_i = a_i + (b_1 X_{1i} + \ldots + b_k X_{ki}) A_i + e_i^*, \text{ and}$$

$$[6] \qquad Y_i = a_i + (b_1 X_{1i} + \ldots + b_k X_{ki}) A_i + v_i + e_i^*$$

where e_i^* consists of e_i and another "error" in the growth <u>rate</u> modeling. Notice that the covariate terms here are all interaction-with-age-at-testing variables. This results directly from the assumption that growth <u>rates</u> rather than intercepts (level of scoring) vary with covariates.[8]

Value-added models are particularly useful in estimating intervention effects on development with a pre- and posttest research design. Pretest data typically are used to estimate average growth curves; deviations from posttest scores predicted by the pretest fitting, then, consist of a number of "error" terms (relative to the quantity of interest, and the value added by the treatment) and the value added. Under appropriate assumptions (e.g., the modeled "errors" have zero mean, are independent of each other, and are independent of the variables fitted in the pretest analysis), the mean of these differences is an unbiased estimator of the average value added.[9]

Although the value-added approach to estimating treatment effects has been used with pre- and posttest designs, typically, it can be used in a posttest only design where the group assignments have been randomly made. Equations [5] and [6] hold for the control and treatment groups, respectively. Of importance here is the assumption that the control and treatment random parameters are sampled from identical populations; random assignment to groups lends strong credibility to this assumption. Both equations can be rewritten in the form of equation [4]. The problem for estimating m_v, the mean value added, is to use the control group data to estimate the mean (or regression) parameters (m_a, b_k, $k = 1, \ldots, K$). Once these estimates are available, an estimate of the quantity ($v_i + e_i^*$) can be obtained for each individual in the treatment group; under the assumptions about e_i^* mentioned earlier, the mean value of these estimates is an unbiased estimator of m_v. When these conditions are met (samples drawn and assigned to groups at random, and the modeled "errors" independent of covariates and the value added), this statistic is easily tested, for any given null hypothesis: a simple t-test is justified, if the sample is of sufficient size to mitigate concerns about nonnormality. For small samples, more robust tests of resistant estimates may be more appropriate.

B.3 Models Employed in Bayley Analysis

The CFRP research design includes elements of experimental and quasi-experimental design. Clearly, the populations available at each of the six impact study sites are "naturally occurring" populations; that is, assignment of families to these sites is (for all practical purposes) uncontrolled. Within each site, however, a sample of eligible families was recruited during the summer 1978; from each pool, families were assigned randomly to CFRP and control/comparison groups. As of the spring 1979 data collection, attrition, while of serious magnitude, did not appear to be differentially selective across control/comparison and CFRP groups (Nauta, et al., 1980, Appendix A). Attrition from spring 1979 to the time of Bayley data collection was much lower than that experienced during the first program year of the evaluation. At this point, the two groups within each site still seem to be sampled from identical initial populations.

CFRP intent concerning child development is clear: by working through and with children's families, CFRP involvement is expected to promote children's development. All CFRP children are expected to benefit. Nevertheless, the mechanism or form that this intervention benefit is expected to take is not clear. Whether the benefit expected consists of a uniform boost in developmental measures and skills or a differential boost according to child and family background characteristics is unknown, a priori. Arguments can be made for either type of expected program benefit, particularly when the relative age and SES homogeneity of the evaluation sample is considered. Models of both types (traditional ANCOVA and value-added) were employed in exploring program impacts on Bayley scores.

The six sites still pose problems for analysis, however. In addition to the potential problem of site-by-CFRP interactions, it is quite possible that regression surfaces are not parallel in the separate sites. (This is a problem for any value-added modeling as well as for an ANCOVA.) Although a variety of interactions with sites can be modeled and tested, it is inconceivable that all possible interactions with site could be considered in any useful way. Two separate sets of analyses were conducted for both types of modeling: global analyses (including data from all sites) and within-site analyses. Although within-site analyses will have far less statistical power (due to small sample size), they can offer some protection against inadvertent gross misspecification of global models.

Prior to estimating (and testing) "final" models of impact, much exploration was undertaken: outliers were screened and many data transformations were employed. The natural logarithms of per capita income and household size were taken, for instance. Dummy variables were coded (-1, 1) in order to mitigate, to the extent possible, problems of multicollinearity in estimating models with interactions between these dummy variables and continuous measures. Age at testing, when also used in interaction with dummy variables, was adjusted by subtracting an arbitrary constant, so that the interaction variables would be continuous. [With age coded in months, from 15.00 upward, an age-by-dummy interaction with the dummy coded (-1,1) will be distributed in two distinct clumps, one with values less than or equal to -15, the other with values greater than or equal to 15. Arbitrarily forcing age to span the value of zero remedies this two-clump distribution problem.] Finally, the relationship between age at testing and Bayley scores was examined in detail, primarily among control/comparison group children. Effectively, the age span was from 15 to 23 months, inclusive. Within this age range, the relationship between age and test scores appears to be linear; there is no strong evidence of "floor" or "ceiling" effects. No attempted modeling of that relationship improved it. (A quadratic modeling of test scores on age and age-squared was tried; the natural logarithm and square root transformations of the test scores were also regressed on age alone, with no improvement.) Therefore, a linear relationship of test scores with age was assumed in the models used, throughout.

ANCOVAS: The ANCOVA began with regressions of test scores on age, dummy variables for site and group (CFRP, control/comparison), and site-by-group interactions dummy variables. (As noted in Chapter 3, St. Petersburg was eliminated from the ANCOVAs and growth model estimation in the value-added analyses.) Outliers were located (again) by examination of residual scatterplots with age, and frequency listings within sites and groups. After a preliminary age and site model was fitted (there were no CFRP effects), other sets of possible covariates were considered. The first set consisted of SES, family, and child characteristics; the second included a number of attitudinal measures derived from earlier parent interviews (see Nauta, et al., 1980). Finally, site and CFRP interactions with covariates were explored. Within-site ANCOVAs considered only age, the first set of covariates mentioned above, a CFRP effect, and then a CFRP-by-age interaction.

Value-added: An average growth model in the absence of treatment was estimated first (using only control/comparison group children). Initially, scores were regressed on age, age-by-site interactions, and site dummy variables, and a preliminary growth model constructed. Next, two sets of variables were considered, one containing age-by-SES, -family, and -child characteristics, the second with age-by-attitudinal measures. Site-by-age-by-covariate interactions were explored next. A final growth model was then estimated, and used in the CFRP group to predict Bayley scores. The difference between observed and predicted scores contains the individual value added estimate plus "errors"; the mean difference estimates the average value added by CFRP participation. Due to the relatively small control/comparison group sizes within sites, within-site growth modeling considered age at testing only.

Notes

1. There are a number of other issues to be considered in judging the practical importance of conclusions drawn from such tests, of course, in addition to consideration of the statistical significance obtained. For instance: Were the treatments actually implemented? Does the research design (including sample sizes) have sufficient power to detect effects of any practical significance? Is the design <u>too</u> powerful, e.g., is (are) the effect(s) detected too small to be of any practical importance (even though significant)?

2. Univariate and bivariate outliers are located easily, typically using visual aids of some kind (histograms, scatterplots). Multivariate outliers are particularly problematic for statistical estimation and testing, but often very difficult to detect (see Gnanadesikan, 1977; Hoaglin and Welsch, 1978).

3. The value of random assignment to treatment and control groups, of course, is to lend strong credibility to the assumption that "errors" in the model are distributed independently of group assignment or membership. But even with random assignment, particularly with small samples, "errors" <u>may</u> be distributed differentially in the two groups.

4. It should be noted that the historical and theoretical development of ANCOVA was based on its value in increasing statistical power in randomized experiments, <u>not</u> on its use as an "adjustment" mechanism for eliminating bias due to initial nonequivalence of the groups. Indeed, blind, unthinking use of ANCOVA models has come in for serious and warranted criticism (Elashoff, 1969; Cronbach, et al, 1976). Nonetheless, the subtle, complex, and serious problems of initial group nonequivalence do not necessarily pose insurmountable problems for analysis (Rubin, 1974, 1977, and 1979).

5. If group assignments were made randomly, or if the covariate(s) is (are) distributed identically within the groups, an argument for the "reasonableness" of the ANCOVA effect estimator can be made. Simply, it is the "average" program effect across all values of the covariate. If the covariate(s) is (are) distributed differentially within the two groups, and this is true necessarily with any "nonequivalent control group" design, the ANCOVA estimator fails even in this sense. Furthermore, where "heterogeneity of regression" is found--whether intended or not, and regardless of degree of initial group nonequivalence--the non-parallelism of

Notes (continued)

5. (continued)

regression surfaces will almost certainly have important policy implications. The effective use of limited resources in implementing a program that appears to affect people differentially requires selective focusing in the distribution and/or use of those resources. Further "adjustments" in estimating treatment effects are required (Rubin, 1974 and 1977). [A special case of this "heterogeneity of regression" problem can be found in aptitude-by-treatment interaction (ATI) applications in educational research.]

6. Applications of within-group regression slope models of treatment outcomes can be found under the rubric of "multi-level" analyses in educational evaluation. Complex and subtle issues of statistical testing and causal attribution abound here (Bache, 1980).

7. If this confounding of unmeasured variables with group membership is serious or extensive, or if it is suspected or known but the confounding variables have not been measured, it is impossible to attribute cause to the program intervention. (Bache and Nauta, 1979, provide a very clear example of just this problem in a Home Start Followup evaluation.) With moderate confounding, an understanding of the selection process which renders the groups nonequivalent, and if the relevant selection variables are measured, estimates of treatment effects--conditional on values of the covariate(s)--with reasonable attribution of cause to the intervention can be developed (Rubin, 1977).

8. This does not <u>prohibit</u> models where covariates affect level of scoring on a developmental test rather than rates of growth. In contrast to ANCOVA models, however, the interaction-with-age variables assume priority over simple covariates.

9. Testing this estimate formally is not so straightforward, and this estimator is inefficient (see Bryk, Strenio, and Weisberg, 1980).

References

Bache, W.L., III, 1980. The indeterminacy of treatment effects when pretreatment regression slopes are heterogeneous. Presented at the annual meeting of the American Educational Research Association, Boston, MA, April 1980.

Bache, W.L., III, and Nauta, M.J., 1979. The Home Start Followup Study: Final Report. Cambridge, MA: Abt Associates Inc.

Bryk, A.S., and Weisberg, H.I., 1976a. Value-added analysis: A dynamic approach to the estimation of treatment effects. Journal of Educational Statistics, 1, 127-155.

Bryk, A.S., and Weisberg, H.I., 1976b. Use of the non-equivalent control group design when subjects are growing. Mimeo from the Huron Institute, Cambridge, MA.

Bryk, A.S., Strenio, J.F., and Weisberg, H.I., 1980. A method for estimating treatment effects when individuals are growing. Journal of Educational Statistics, 5, 5-34.

Campbell, D.T., and Stanley, J.C., 1963. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally College Publishing Company.

Cronbach, L.J., Rogosa, D., Floden, R.E., and Price, G., 1976. Analysis of Covariance--angel of salvation or temptress and deluder? Occasional Paper, Stanford Evaluation Consortium, Stanford University, Stanford, California.

Elashoff, J.D., 1969. Analysis of covariance: a delicate instrument. American Educational Research Journal, 6, 383-401.

Gnanadesikan, R., 1977. Methods for Statistical Analysis of Multivariate Observations. New York: John Wiley & Sons.

Hoaglin, D.C., and Welsch, R.E., 1978. The hat matrix in regression and ANOVA. American Statistician, 32, 17-22.

Mosteller, F., and Tukey, J.W., 1977. Data Analysis and Regression: A Second Course in Statistics. Reading, MA: Addison-Wesley Publishing Company.

Nauta, M., Brush, L., Johnson, L., Affholter, D., and Hewett, K., 1980. Evaluation of the Child and Family Resource Program: Phase II Report (Volume I:Research Report). Cambridge, MA: Abt Associates Inc.

Rubin, D.B., 1979. Using multivariate matched sampling and regression adjustment to control bias in observational studies. Journal of the American Statistical Association, 74, 318-328.

References (continued)

Rubin, D.B., 1977. Assignment to treatment group on the basis of covariate. Journal of Educational Statistics, 2, 1-26.

Rubin, D.B., 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology, 66, 688-701.

Theil, M., 1971. Principles of Econometrics. New York: John Wiley & Sons.

Appendix C

IMPACT ANALYSES

As noted in Chapter 3, two separate sets of analyses of CFRP impact on Bayley scores were conducted, employing standard ANCOVAs and a value-added approach. The theoretical assumptions behind each are elaborated in Appendix B. The analyses themselves are described here.

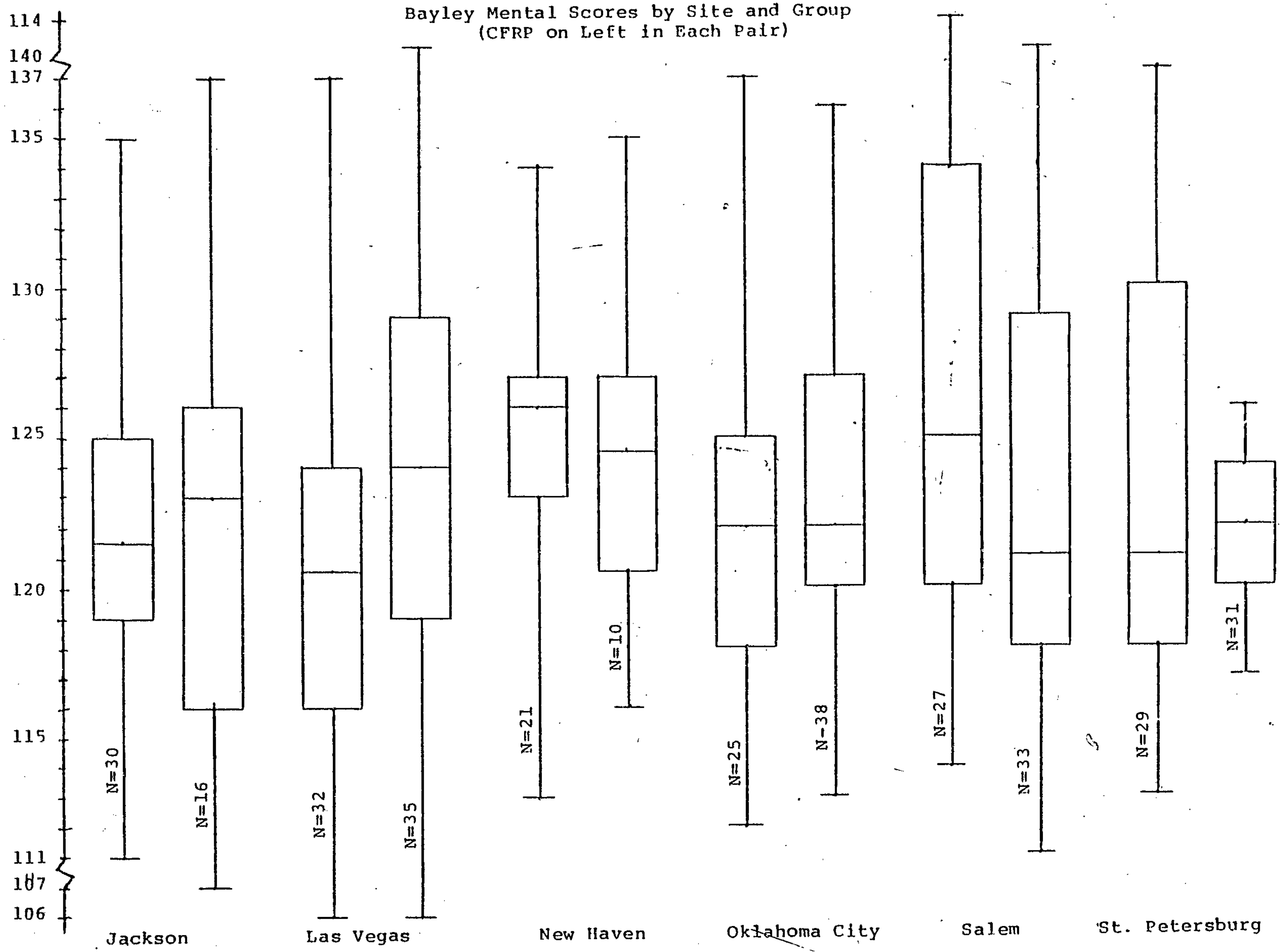
C.1 Mental Score Analyses

Analysis began with examination of the relationship between age and Mental Score. Scatterplots within each site-by-group classification were examined for outliers. As discussed in Chapter 3, the data for the St. Petersburg control/comparison group were anomalous; instead of following the typical pattern of positive correlation between age and Mental Score, the plot in this group was flat ($\underline{r}$ = - 01). Therefore, St. Petersburg was excluded from ANCOVA tests of CFRP effects and from the growth model estimation prior to making global value-added tests. (The St. Petersburg CFRP group was included, however, in the estimation and testing of global value added.) Figure C-1 displays Mental Scores by site and group; the St. Petersburg anomaly is quite striking.

Within the age range tested, the age/Mental Score relationship seems to be linear. There is no evidence of floor or ceiling effects, or of curvilinearity in the developmental plots. A number of alternatives were examined: Mental Score was regressed on age and age squared; the natural logarithm of Mental Score was taken--an appropriate transformation if the growth curve is bowed downward, as would be expected with floor effects, or with a lower asymptote--and tried as an alternative to the raw score.

Figure C-1

Bayley Mental Scores by Site and Group
(CFRP on Left in Each Pair)



A linear relationship between age and Mental Score was assumed throughout the analyses reported below.

ANCOVAs: Two simple ANCOVAs (site, CFRP, with age as a covariate) were conducted. The first (Table C-1) included all sites except St. Petersburg, and showed no CFRP effect. In the second (Table C-2), both St. Petersburg and New Haven were excluded, because of the apparent tester problems at these two sites (see Appendix A); again, there was no CFRP effect.

Table C-1

ANCOVA of Mental Score[a]

Source	SS	df	MS	F	p
Age	6489.06	1	6489.060	198.30	.000
Main:					
Site	184.29	4	46.07	1.41	.232
CFRP	22.87	1	22.87	0.70	.404
Site x CFRP	106.07	4	26.52	0.81	.502
Residual	8409.82	257	32.72		
TOTAL	15743.14	267	58.96		

(b_{age} = 1.80; this represents a growth rate of 1.8 points on Mental Score per month of age)

[a] St. Petersburg excluded.

Table C-2

ANCOVA of Mental Score[b]

Source	SS	df	MS	F	p
Age	6041.45	1	6041.45	210.53	.000
Main:					
Site	167.48	3	55.83	1.94	.123
CFRP	34.64	1	34.64	1.21	.273
Site x CFRP	106.01	3	35.34	1.23	.299
Residual	6427.96	224	28.70		
TOTAL	12806.32	232	55.20		

(b_{age} = 2.23)

[b] New Haven and St. Petersburg excluded.

The site and CFRP measures were recast as dummy variables for further exploration of ANCOVA models (New Haven included). As expected, no dummy variables for sites are significant, but there appears to be a small Salem-by-age interaction: without considering other covariates, the estimated growth rate for all other sites is about 2.00 points per month, while in Salem it seems to be about 2.64 points per month. (A two-tailed test of the interaction coefficient is significant at .06.) A number of covariates were considered next, in four sets:

- Child characteristics--sex, race, whether the oldest child, and whether "at risk" due to circumstances at birth (low birth weight or other physical problems).
- SES--the natural logarithm of per capita income, an income source construct, whether or not the child's mother was employed, and whether she had finished high school at CFRP entry.
- Family characteristics--the natural logarithm of household size, and whether the child's mother had ever been married (formally or informally).
- Other "softer" variables (constructed from mother's self-reported items)--the extent of social interaction with friends, the extent of social interaction in organized groups (church, political clubs, etc.), a measure of "hassled" feelings (generally), an assessment of the numbers of "worried" and "pleasing" situations typically faced (intended to measure "negative" and "positive" reinforcement for parenting roles or activities), and two measures of infant temperament (one including primarily reports of physical well-being--e.g., a health problem requiring frequent attention, feeding problems and sleeping problems--and the other containing items of a more social nature--for instance, difficulty in comforting, and doesn't smile or seems disinterested in things going on around him/her).

Finally, site-by-covariate interactions were explored. (At no point during the model-building process were CFRP or CFRP-by-age interactions significant.)

The model finally adopted includes age, the Salem-by-age interaction, two covariates (whether the mother was ever married and extent of social interaction with friends), and five site-by-covariate interactions (Jackson-by-whether mother employed; Las Vegas-by-whether mother employed and Las Vegas-by-log of per capita income; and New Haven-by-sex and New Haven-by-whether the child is the oldest in the family). The estimated model is summarized in Table C-3. Very little explanatory power is added with the inclusion of all variables other than age: R^2 (age alone) = .44; with Salem-by-age, R^2 = .45; in Table C-3, R^2 = .46 (all R^2 "adjusted"

Table C-3

Global ANCOVA Model of Mental Score[a]

	b	s.e.	p[b]
Age[c]	2.05	.19	<.001
Salem-by-age[d]	.39	.18	.033
Mother ever married[d]	1.28	.41	.002
Social interaction with friends[d]	.89	.46	.052
Jackson-by-mother employed[d]	.68	.41	.092
Las Vegas-by-mother employed[d]	-.77	.41	.062
Las Vegas-by-ln (per capita income)[d,e]	-.82	.57	.151
New Haven-by-child's sex[d]	.66	.38	.078
New Haven-by-whether oldest child[d]	-.71	.41	.086
Intercept[c]	120.35	R^2=.46	

[a]St. Petersburg excluded; N ranges from 221 to 265 (pair-wise deletion used in estimating regression).

[b]Excepting the test for age, all tests are two-tailed.

[c]Age is age at testing, in months, minus 18 months; the intercept, therefore, is the predicted score at age 18 months.

[d]All dummy variables are coded (-1,1).

[e]The natural logarithm of per capita income ($K) was used.

for the number of covariates in the equation). CFRP and CFRP-by-age variables simply have no evident effect on Mental Score; correlations and partial correlations appear in Table C-4.

The ANCOVA model presumes parallel regression surfaces ("homogeneity of regression") within each of the site-by-CFRP groups, excepting those violations detectable as interactions shown in Table C-3. Given the number of site-by-covariate interactions detected, as well as the known differences between sites, separate within-site ANCOVAs were explored. St. Petersburg was excluded, again. Excepting Salem, the models adopted are very simple and similar (Table C-5). Age, as expected, is the most important covariate in all sites--even though it does not "explain" much in New Haven. Whether the tested child is the oldest child (oldest children score higher) appears to be somewhat important in Las Vegas and Oklahoma City. The extent of the mother's social interaction with friends enters in two sites (more interaction, higher scores), Oklahoma City and Salem. Three more covariates are included in the Salem model: child's sex (girls score higher), the income source construct (families who rely more on earned income have children who score higher), and the "high risk" measure (children at risk score lower). In all cases, CFRP and the age-by-CFRP measure are unimportant; there is no evidence of a positive influence on Bayley Mental Scores.

Table C-4

Correlations with Mental Score[a]

	CFRP[b]	Age-by-CFRP
Zero order	-.02	-.03
Controlling age	-.05	-.03
Controlling age and Salem-by-age	-.06	-.03
Also controlling covariates	-.05	-.01
Also controlling site-by-covariates	-.06	-.00

[a]St. Petersburg excluded; total N=265.

[b]CFRP is coded (-1,1); -1 is the control/comparison group.

Table C-5

Within-Site ANCOVAs, Mental Score[a]
(standard errors in parentheses)

	Jackson	Las Vegas	New Haven	Oklahoma City	Salem
Age[b]	1.69 (0.41)	2.51 (0.45)	0.86 (0.52)	1.56 (0.26)	2.99 (0.27)
Oldest child[c]	--	3.12 (2.01)	--	0.71 (0.56)	--
Social interaction with friends[c]	--	--	--	1.01 (0.64)	1.82 (0.80)
Sex[c]	--	--	--	--	-0.89 (0.61)
Income source construct	--	--	--	--	1.19 (0.75)
High risk[c]	--	--	--	--	-1.05 (0.81)
Intercept[b]	122.17	119.98	123.68	119.88	119.99
R^2	.26	.47	.06	.47	.68
Partial correlations with Mental Score					
CFRP	-.10	-.18	-.12	-.20	.23
CFRP-by-age	-.16	-.11	-.37	-.10	.08

[a] St. Petersburg excluded.

[b] Age is age at testing less 18 months; the intercept is the predicted score at age 18 months.

[c] All dummy variables are coded (-1,1).

Value-added: The first step in a value-added modeling of Mental Score was to estimate a growth model among control/comparison scores (St. Petersburg excluded). Table C-6 gives descriptive statistics on the variables included in the model chosen; Table C-7 contains the corresponding correlations; Table C-8 summarizes the model fitted. In settling upon this model, groups of variables were considered in the following order: age, site, and age-by-site interactions; age interactions with SES, family, and child characteristics; and age interactions with parental behavior and attitudinal measures. The widespread searching strategy leading up to adoption of this model capitalizes on sampling

Table C-6

Descriptive Statistics, Variables in the Global Mental Score Growth Model[a]

	Mean	SD	N[b]
Mental Score	123.60	7.58	128
Age[c]	0.11	2.32	130
Age interactions			
Oklahoma City[d]	0.01	2.28	129
High risk[d]	0.19	2.32	129
Per capita income[e]	-0.02	1.70	100
Ever married[d]	0.46	2.24	127
Black[d]	-0.53	2.27	129
Salem-by-oldest child[d]	0.17	2.32	129

[a]Control/comparison sample only; St. Petersburg excluded.

[b]Missing data primarily are due to missing covariate information, although a few extreme age-by-covariate values were recoded as missing.

[c]Age is at testing less 18 months.

[d]Dummy covariates are coded (-1,1).

[e]The natural logarithm of per capita income ($K) was used.

Table C-7

Correlations Among Variables in the Global Mental Score Growth Model[a]

	Mental Score	Age[b]	Age interactions				
			Oklahoma City[c]	High risk[c]	Per capita income[d]	Ever married[c]	Black[c]
Age[b]	.69						
Age interactions							
Oklahoma City[c]	-.47	-.56					
High risk[c]	-.40	-.44	.05				
Per capita income[d]	.37	.64	.57	-.21			
Ever married[c]	.25	.27	-.12	-.15	.36		
Black[c]	-.16	-.26	.38	.14	-.25	-.50	
Salem-by-oldest child[c]	-.07	-.06	.38	.08	-.02	.00	-.14

[a]Control/comparison group only, St. Petersburg excluded.

[b]Age is at testing minus 18 months.

[c]Dummy covariates are coded (-1,1).

[d]The natural logarithm of per capita income ($K) was used.

Table C-8

Global Mental Score Growth Model[a]

	b	s.e.	p[b]
Age[c]	1.76	.352	<.001
Age interactions			
Oklahoma City[d]	-1.68	.411	<.001
High risk[d]	-0.72	.268	.009
Per capita income[e]	-1.47	.460	.002
Ever married[d]	0.94	.304	.003
Black[d]	0.99	.326	.003
Salem-by-oldest child[d]	-.69	.286	.018
Intercept[c]	123.51		
R^2	.55		

[a]Control/comparison group only, St. Petersburg excluded.

[b]Excepting age, all tests are two-tailed.

[c]Age is age at testing minus 18 months; the intercept is the predicted score at 18 months.

[d]Dummy covariates are coded (-1,1).

[e]The natural logarithm of per capita income ($K) was used.

(or random) errors in selecting variables for inclusion, so any interpretation of the model must be viewed skeptically. Nonetheless, the model estimated seems reasonable. With all other variables "controlled," and within the age range tested, in addition to the expected age regression (or average growth), it suggests:

- The typical age regression (or growth rate) in Oklahoma is flatter, by just over three points per month, than for children at the other sites.
- The typical age regression (or growth rate) among children at risk is flatter, by about one and one-half points per month, than for other children.

- As per capita income increases, the age regression (or growth rate) decreases.
- The typical age regression (or growth rate) of children in families where the mother has never married is flatter, by just over two points per month, than for children in other families.
- The typical age regression (or growth rate) of black children (in this sample) is steeper, by just under two points per month, than for non-black children.
- The typical age regression (or growth rate) among oldest children in Salem is steeper, by just under one and one-half points per month, than for other children.

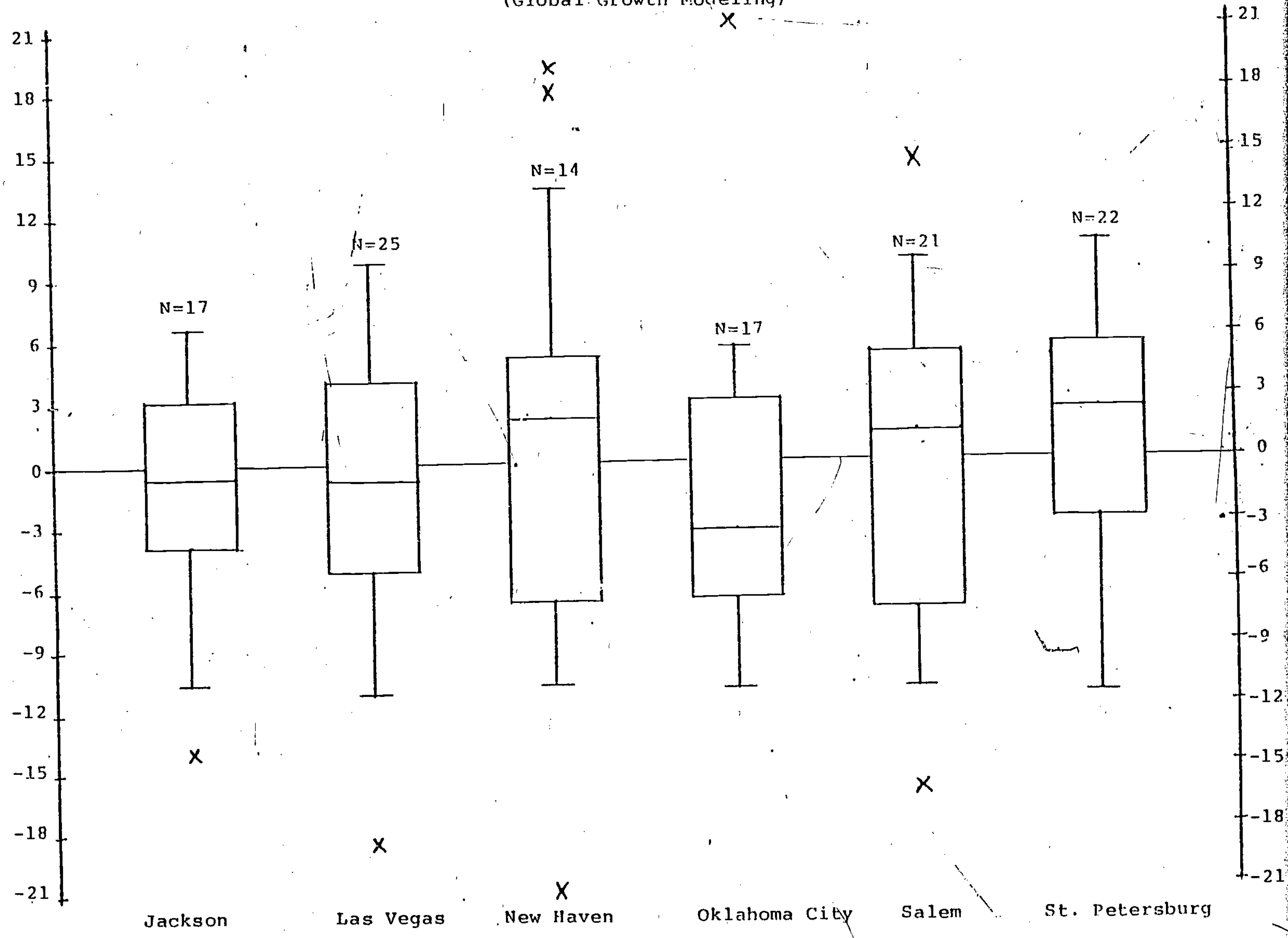
There is a strong caveat here: the regression coefficients are "partial" coefficients--that is, the estimated effects with all other variables in the equation held "constant," statistically. For many variables, this "controlling" or "holding constant" has little relevance to what actually happens, as a glance at the correlations in Table C-7 will show. Consider the age-by-Oklahoma City and age-by-race effects: one boosts the estimated growth rate, while the other depresses it, other things being equal. But the Oklahoma City sample is predominantly, although not wholly, black; in actuality, then, these two estimated "effects" tend to "cancel" each other. Also, consider the age-by-per capita income effect. In Table C-7, this variable has a fair-sized positive correlation with Mental Score (.36). Yet, with other variables "controlled," the "effect" is in the opposite direction. (Indeed, it is precisely the substantial correlations shown in Table C-7 that lead to interpretive difficulties. If two variables are relatively uncorrelated, then "controlling" for one does not alter an interpretation of what the other measures, in effect. When two variables are relatively highly correlated, however, it is difficult to imagine just what meaning can or should be attributed to either when the other is "controlled.")

Given the growth model estimated in the control/comparison group, Mental Scores for the CFRP group were predicted, and the (observed-less-predicted) differences taken. These estimate the value added by CFRP participation, and the mean of these differences is an unbiased estimator of the mean CFRP effect. The mean of these differences is -.248, indicating no global CFRP impact (SD = 7.421, N = 116; p >.50). A one-way ANCOVA by site fails to reject the null hypothesis of equal means by site. There are no outliers evident in the distribution of these differences; all but two fall within two and one-half standard deviations of zero (as 99% of any set of observations sampled from a normal distribution with zero mean should be), and these two are just outside that interval (one positive, one negative). Figure C-2 summarizes the distribution of value-added estimates by site. Nowhere is there any indication of an important (or statistically significant) CFRP effect.

For four sites, value-added analyses were repeated within site. St. Petersburg, again, could not be used. New Haven's sample of only 12 control/comparison cases cannot support estimation of any growth model. (Indeed, the utility of value-added testing within sites is limited severely by small sample sizes.) Summaries of the four growth models fitted are given in Table C-9. Value-added estimates were generated within the CFRP groups at these sites, and outliers were screened using an arbitrary criterion of outside two (estimated) standard deviations of the sample mean. Results are summarized in Table C-10.

Figure C-2

Value Added, Bayley Mental Scores, by Site[a]
(Global Growth Modeling)



[a]Each "X" represents an outlier.

Table C-9

Within-Site Mental Score Growth Models[a]

	Jackson	Las Vegas	Okla-homa City	Salem
Age[b]	2.61	2.74	1.66	2.71
Intercept[b]	123.95	123.45	122.48	123.73
N	16	25	36	33
R^2	.48	.53	.44	.84

[a] Control/comparison groups only, St. Petersburg and New Haven excluded; see text. Within-site models include only age, that is, they do <u>not</u> model the average growth rate as a function of covariates.

[b] Age is age at testing minus 18 months; the intercept is the predicted score at age 18 months.

Table C-10

Within-Site Value-Added Summaries

	Jackson	Las Vegas	Okla-homa City	Salem
Mean	-1.78	-0.71	-1.86	2.18[c]
SD	5.08	4.71	3.03	4.51
Median of Walsh averages[d]	-1.69	-0.69	-1.82	2.49
Sample median	-1.30	0.06	-2.29	3.03
N	29	29	23	27
N omitted[e]	1	1	1	1

[c] The mean value-added in Salem is significant at .04 (one-tailed test, with simultaneous control for Type I error rates across four tests).

[d] A Walsh average is the average of any pair of observations. The median of Walsh averages is the median of all possible Walsh averages; it is a more efficient estimator of a population median than is the sample median.

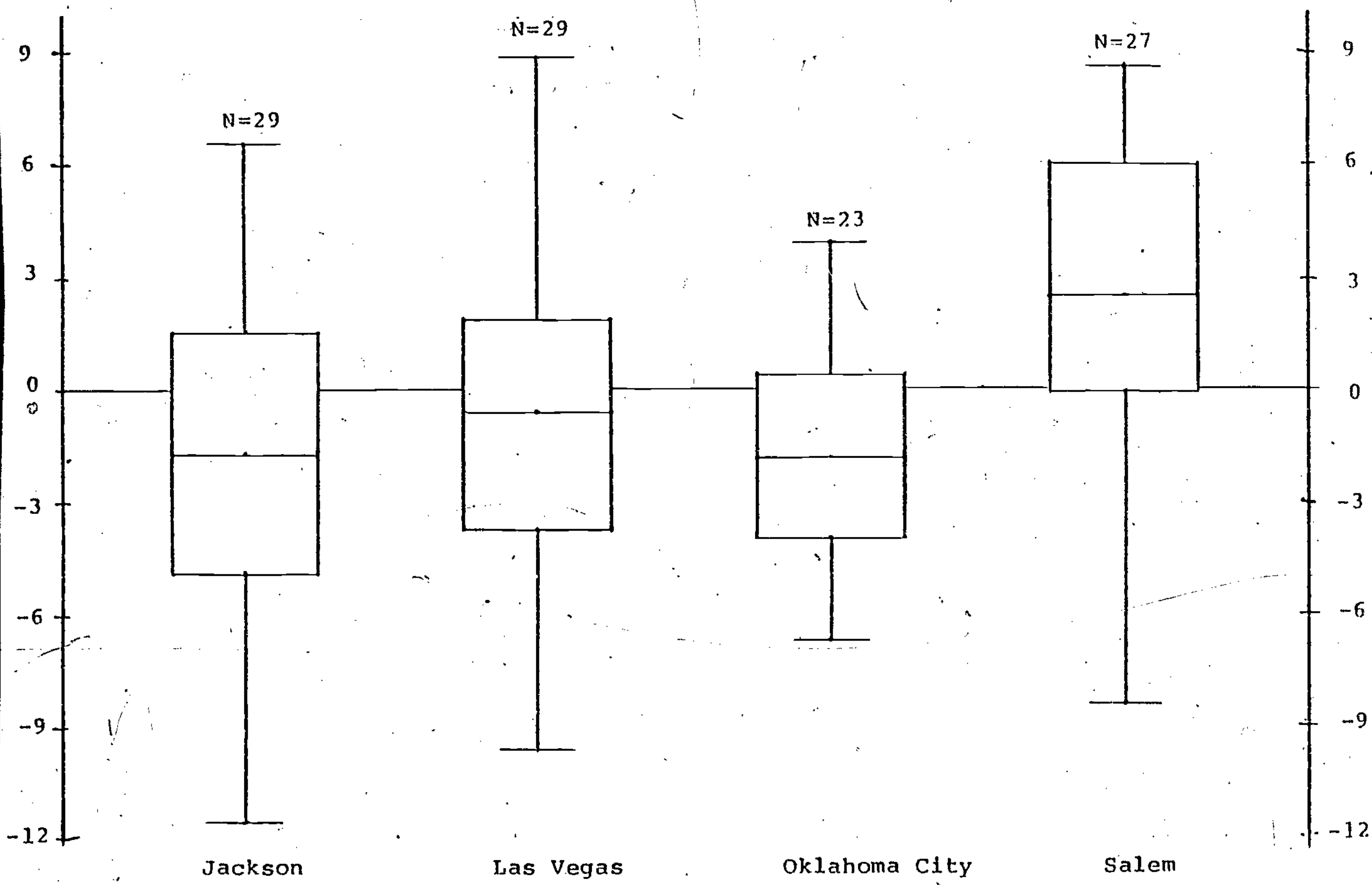
[e] Outliers were screened, using a two-SD rule; N omitted shows how many.

It appears that CFRP in Salem has a statistically significant positive effect on Bayley mental scores: a one-tailed t-test, controlling the Type I error rate over four simultaneous tests, reaches a significance level of less than .04. Nevertheless, this "effect" must be viewed with considerable skepticism: the small-sample growth models can hardly be adequate for explaining mental development. Figure C-3 summarizes within-site value-added results.

In summary, with only a tenuous hint at positive CFRP impact in Salem, there is no evidence that CFRP involvement has enhanced the mental development of the focal children in the evaluation study families, as measured by the Bayley Scales of Mental Development.

Figure C-3

Value Added, Bayley Mental Scores, by Site
(Within-Site Growth Modeling)



C.2 Physical Score Analyses

The analysis began with exploration of item-by-item scoring patterns. As noted in Chapter 3, only ten items from the Physical Development Scale had adequate variation in pass/fail rates to be useful for analysis; these items served as a basis for computation of Physical Score. This measure represents the proportion of items passed for each case with valid values on eight or more of these items. All analyses were performed on this Physical Score.

ANCOVAs: An initial ANCOVA, with age as the only covariate, showed no CFRP effects and no site-by-CFRP interaction (Table C-11); only age and site effects were found. Site and CFRP were recast as dummy variables, and a number of stepwise regressions were examined, following the strategy outlined previously for analyzing the Bayley Mental Score. Two site dummy variables are significant (Jackson and Las Vegas children score lower, typically, than children at other sites), as are four covariates (whether the child's mother is employed, the child's birth weight, whether the child is at risk, and the natural logarithm of per capita income). Unlike the Mental Score ANCOVA, no site-by-covariate interactions contribute anything to the Physical Score model. The model estimated is summarized in Table C-12. Correlations of CFRP and CFRP-by-age with Physical Score are shown in Table C-13. As in the Mental Score analysis, there is no evidence of CFRP impact.

11

Table C-11

Global ANCOVA, Physical Score[a]

Source	SS	df	MS	F	p
Age	2.760	1	2.760	66.75	.000
Main:					
Site	1.139	4	0.285	6.88	.000
CFRP	0.038	1	0.038	0.92	.339
Site-by-CFRP	0.023	4	0.006	0.14	.969
Residual	10.214	247	0.041		
TOTAL	14.784	257	0.058		

(b_{age} = 0.045)

[a] St. Petersburg excluded.

Table C-12

Global ANCOVA Model of Physical Score[b]
(coefficients and their standard errors are in 10^2)

	b	s.e.	p[c]
Age[d]	4.22	0.60	<.001
Jackson[e]	-8.67	1.91	<.001
Las Vegas[e]	-5.10	1.77	.008
Mother employed[e]	4.84	1.62	.006
Birth weight	-2.60	1.25	.075
High risk[e]	-3.00	1.77	.180
Per capita income[f]	3.65	2.45	.273
Intercept[d]	59.67	R^2=.33	

[b] St. Petersburg excluded.

[c] Excepting the age coefficient, all tests are two-tailed.

[d] Age is age at testing minus 18 months; the intercept is the score predicted at age 18 months.

[e] Dummy covariates are coded (-1,1).

[f] The natural logarithm of annual per capita income ($K) was used.

Table C-13

Correlations with Physical Score[a]

	CFRP	Age-by-CFRP
Zero order	-.06	-.07
Controlling age	-.09	-.08
Also controlling site	-.05	-.09
Also controlling covariates	-.02	-.11

[a]St. Petersburg excluded.

As noted above, this approach to statistical modeling casts a wide net in looking for potentially relevant covariates; hence, any interpretation of the models estimated must be viewed skeptically. In a very nontrivial sense, this approach to modeling capitalizes on the vagaries of sampling (and measurement) error in attributing "importance" to relationships discovered through comprehensive search. Nevertheless, the models estimated suggest that a number of characteristics can influence physical development (as measured by Physical Score) substantially: employment, for instance, is "worth" more than two months growth, typically, in this sample. [Since dummy variables are coded (-1,1), the "effect" is twice the regression coefficient. Also, note that per capita income is in natural logarithm units (of thousands of dollars). A change from $1500 to $2000 annual per capita income is a change of about 0.29 units of income modeled, while a change from $2000 to $2500 is about 0.22.] Being "at risk" costs about one and one-half months growth, in this sample.

Again, separate ANCOVAs within sites were explored. Table C-14 summarizes the important findings--or lack thereof in the case of CFRP impact tests. Evidently, CFRP has nc positive impact on Bayley Physical Scores--indeed, the values of partial correlations in Jackson, New Haven, and Salem hint at a <u>negative</u> CFRP effect.

Table C-14

Within-Site ANCOVAs of Physical Score Summarized[a]

(intercept and age coefficients are in 10^2 units)

	Jackson	Las Vegas	New Haven	Oklahoma City	Salem
Intercept	35.84	63.84	52.49	84.22	77.02
Age slope[b]	4.54	6.89	5.94	3.02	3.95
(s.e.)	(1.85)	(1.62)	(2.67)	(1.37)	(1.14)
R^2	.15	.37	.23	.31	.20
N[c]	43	51	30	59	63
Partial correlations with Physical Score:					
CFRP	-.26	-.09	-.10	.03	.09
CFRP-by-age	-.36	-.12	-.23	.02	-.20

[a] St. Petersburg omitted.

[b] The covariates included in each model varied by site; the age coefficient given here is a "partial" slope, "controlling" for those other covariates.

[c] Effective Ns for hypothesis testing varied, due to the use of the pairwise deletion of missing data values. The Ns given here are total sample sizes.

Value-added: The estimation of the value-added model was done in the control/comparison group, excluding the St. Petersburg subsample. (As noted, peculiarities in the mental scores among this group are evident and inexplicable.) Table C-15 gives descriptive statistics on the set of variables included in the growth model estimation; Table C-16 lists their intercorrelations. The initial growth model is given in Table C-17.

Table C-15

Descriptive Statistics, Variables in the Global Physical Score Growth Model[a]

	Mean	SD	N[b]
Physical Score	0.53	0.24	121
Age[c]	0.11	2.32	130
Jackson[d]	-0.73	0.69	132
Las Vegas[d]	-0.61	0.80	132
Age interactions			
Jackson[d]	-0.25	2.31	130
Sex[d]	0.22	2.31	130
Jackson-by-per capita income[d,e]	0.01	1.70	100
Jackson-by-finished high school[d]	-0.08	2.29	127
New Haven-by-finished high school[d]	-0.38	2.26	127
Oklahoma-by-per capita income[d,e]	0.06	1.69	100
Oklahoma-by-household size[d,e]	-0.11	3.56	128

[a] Control/comparison group only, St. Petersburg excluded.

[b] Outliers were screened, but most missing data represent missing rather than extreme information. Pairwise deletion was used in estimating the model.

[c] Age is age at testing minus 18 months.

[d] Dummy covariates are coded (-1, 1).

[e] The natural logarithms of per capita annual income ($K) and household size were used.

Table C-16

Correlations Among Variables in the Global Physical Score Growth Model[a]

	Physical Score	Age[b]	Jackson[c]	Las Vegas[c]	Age interactions: Jackson[c]	Sex[c]	Jackson-by-income[c,d]	Jackson-by-H.S.[c]	New Haven by-H.S.[c]	Oklahoma-by-income[c,d]
Age[b]	.58									
Jackson[c]	-.23	-.11								
Las Vegas[c]	-.16	-.04	-.20							
Age interactions										
Jackson[c]	-.34	-.78	-.05	.07						
Sex[c]	-.17	-.02	-.09	.08	.01					
Jackson-by-income[c,d]	-.20	-.43	-.01	.04	.64	-.14				
Jackson-by-H.S.[c]	-.12	-.27	.12	-.07	.27	-.13	.33			
New Haven-by-H.S.[c]	-.25	-.35	.07	-.03	.35	-.19	.20	.64		
Oklahoma-by-income[c,d]	-.13	-.50	.03	-.01	.30	-.24	.37	.20	.35	
Oklahoma-by-household size[c,d]	-.36	-.41	.07	-.01	.22	-.03	.22	.07	.13	.55

[a]Control/comparison group only, St. Petersburg excluded.
[b]Age is age at testing minus 18 months.
[c]Dummy covariates are coded (-1, 1).
[d]The natural logarithms of per capita annual income ($K) and household size were used.

Table C-17

Global Physical Score Growth Model[a]
(coefficients and their standard errors are in 10^2 units)

	b	s.e.	p[b]
Age[c]	9.47	1.52	<.001
Jackson[d]	-4.95	2.75	.145
Las Vegas[d]	-4.09	2.25	.139
Age interactions			
Jackson[d]	5.70	1.64	.001
Sex[d]	-1.10	0.81	.350[e]
Jackson-by-income[d,f]	-4.11	1.56	.017
Jackson-by-H.S[d]	2.28	1.09	.073
New Haven-by-H.S.[d]	-3.67	1.15	.003
Oklahoma-by-income[d,f]	6.48	1.52	<.001
Oklahoma-by-household size[d,f]	-1.68	0.61	.011
Intercept	45.06	R^2=.52	

[a] Control/comparison group, St. Petersburg excluded.

[b] Excepting the coefficient for age, all tests are two-tailed.

[c] Age is age at testing minus 18 months.

[d] Dummy covariates are coded (-1,1).

[e] The age-by-sex interaction was significant at <.15 until the various age-by-site-by-covariate interactions were entered.

[f] The natural logarithms of per capita annual income ($K) and household size were used.

There are a number of interesting points to be made, with the caveat made in discussing the Mental Score growth model. Site effects for Jackson and Las Vegas are still evident, but the Jackson-by-age interaction acts to "offset" the negative Jackson effect (that is, scores in Jackson, typically, are below those in other sites except Las Vegas, but the growth rate in Jackson is steeper). All other things "controlled," the estimated growth rate for boys is slower than that for girls. Increasing per capita income in Jackson seems to depress the growth rate, whereas in Oklahoma City per capita income appears to enhance growth, all else "controlled." Whether the child's mother has finished high school has the same flip-flop effect on growth rates in New Haven and Jackson, respectively.

Given this growth model, value-added estimates were derived by subtracting predicted from observed Physical Scores in the CFRP group. A negative value was obtained: -0.019, standard error = .019, $p > .84$. A one-way analysis of variance of these differences showed no differences by site; a breakdown by site is presented in Table C-18. It appears that CFRP involvement has had no positive average effect on physical development as measured by Physical Score.

Table C-18

Value-Added Summaries, Physical Score[a]

	Mean	SD	N
Jackson	-.015	.279	22
Las Vegas	-.020	.181	28
New Haven	-.023	.235	18
Oklahoma	-.032	.200	16
St. Petersburg	-.075	.159	24
Salem	-.043	.246	26

[a]These are global value-added estimates broken down by site.

As with the ANCOVAs, separate analyses were conducted within sites; New Haven was dropped from these analyses due to small samples. The growth models estimated are given in Table C-19.

Value-added estimates at these four sites are summarized in Table C-20. Again, three of the four means are negative, as are three of the four medians. There is no evidence here of positive CFRP effects on Physical Score values, and a suggestion of negative effects. (Note that the only positive estimate is in Salem, although it is not at all significant.)

Table C-19

Within-Site Physical Score Growth Models[a]

(coefficients in 10^2 units)

	Jackson	Las Vegas	Okla-homa City	Salem
Age[b]	10.08	5.43	3.39	4.90
Intercept	46.92	41.31	57.36	57.49
N	16	23	36	34
R^2	.62	.22	.10	.39

[a]Control/comparison groups, St. Petersburg and New Haven excluded.

[b]Age is age at testing minus 18 months.

Table C-20

Within-Site Value-Added Physical Score[c]

(tabled entries, except N, are in 10^2 units)

	Jackson	Las Vegas	Okla-homa City	Salem
Mean	-12.8	-2.3	-3.6	1.74
SD	22.2	15.8	17.2	22.2
Median of Walsh averages[d]	-12.0	-1.9	-3.9	2.5
N[e]	26	29	22	29
N omitted[e]	1	0	1	1

[c]St. Petersburg and New Haven excluded.

[d]A Walsh average is the average of any pair of observations. The median of Walsh averages is the median of all possible Walsh averages; it is a more efficient estimator of a population median than the sample median.

[e]Outliers were screened. N is tested sample size; N omitted shows how many outliers were screened.

Appendix D

PROGRAM PARTICIPATION AND DEVELOPMENT

D.1 Measuring Participation

Program participation is described generally elsewhere in this report (Chapter 2). These measures consist of counts--of the number of home visits, number of parent education sessions, and number of infant-toddler sessions attended. (For purposes of exploring relationships between participation and Bayley scores, participation counts continued only to the date of testing.) The sample size is 151, with participation data missing for five families. (As reported in Chapter 2, New Haven was excluded from all analyses of program participation.) The number of home visits ranged from no visits (two families) to 53 visits (one family), with a median of 14 visits. Twenty-five percent of the sample had 8 or fewer visits; twenty-five percent had 23 or more visits. In contrast, 65 families had no reported attendance at infant-toddler (I/T) sessions, and 78 families had no reported attendance at parent education sessions. The median number of sessions attended is 4 parent education and 6 I/T; the 25th and 75th percentiles are (1, 9) and (3, 11), respectively.

Average participation rates were obtained simply by dividing the total participation measures by the number of months over which that form of participation occurred; the length of time varies, but for most families the time span is either 13 or 18 months. Average home visit rates ranged from 0.0 to 3.9 per month, with a mean of 1.4 per month (median is 1.1), with lower and upper quartiles at 0.8 and 1.7, respectively. Among those who attended parent education sessions (N = 66), participation ranged from 0.2 to 1.7 per month, with a median of 1.0 and lower and upper quartiles at (0.6, 1.0). I/T session participation rates (N = 68) ranged from 0.1 to 4.5 sessions per month, with a median of 1.0 and quartile "hinges" at (0.8, 1.6).

Table D-1 lists selected inter-item correlations. Clearly, total and rate of home visit participation are highly confounded; only the rate was used analytically. Also, total I/T and parent education session attendance are fairly highly correlated, so a composite total center participation measure was created; because it was highly skewed, its natural logarithmic transformation was taken. Four continuous participation measures, then, were used: home visit, I/T session, and parent education session rates, and the natural log of a total center participation construct. Figures D-1 through D-7 give stem-and-leaf displays of and descriptive statistics for these measures. (These are very much like histograms, and can be read in much the same way. Each data point is represented; the value of any point may be read by taking the stem value and adding the leaf decimal, which is rounded. The two largest values in Figure D-1, for example, are 3.8 and 3.9.)

Table D-1

Spearman Correlations, Participation Measures

Total Amount of Participation Measures	Home Visits	I/T Sessions[a]
I/T sessions[a]	.55	
Parent ed. sessions[a]	.40	.76
Rate of Participation Measures		
I/T sessions[a]	.39	
Parent ed. sessions[a]	.10	.05
Home visits, total with rate	.86	
I/T sessions, total with rate[a]	.44	
Parent ed. sessions, total with rate[a]	.06	

[a]These variables are defined only for those families with some appropriate center participation.

Figure D-1

Stem-and Leaf Display, Average Number of Home Visits Per Month

(a for leaf 0, 1; b for leaf 2, 3; c for leaf 4, 5; d for leaf 6, 7; e for leaf 8, 9)

N		
(2)	0a	00
(1)	0b	3
(9)	0c	455555555
(21)	0d	666666666666667777777
(19)	0e	8888888889999999999
(20)	1a	00000000000011111111
(12)	1b	222222333333
(15)	1c	444444444455555
(9)	1d	666667777
(5)	1e	88899
(5)	2a	00111
(3)	2b	223
(1)	2c	4
(7)	2d	6666777
(5)	2e	88899
(2)	3a	01
(2)	3b	23
(1)	3c	4
(2)	3d	66
(2)	3e	89
N=143		

Mean = 1.39 Median = 1.14 SD = 0.83

IR[a] = 0.93 .75IR[a] = 0.70 Skewness = 1.09

(kurtosis -3) = 0.55

[a]IR is the interquartile range; in a normally distributed population, .75IR is about one standard deviation.

Figure D-2

Stem-and Leaf Display, Average Number of I/T Sessions Per Month[a]

(a for leaf .00-.24; b for leaf .25-.49; c for leaf .50-.74; d for leaf .75-.99)

N		
(1)	0a	1
(5)	0b	34444
(7)	0c	5556666
(9)	0d	888888899
(19)	1a	0000000000000011111
(6)	ab	344444
(7)	1c	5556677
(2)	1d	89
(5)	2a	00022
(0)	2b	
(0)	2c	
(0)	2d	
(2)	3a	00
(0)	3b	
(1)	3c	5
(0)	3d	
(3)	4a	000
(0)	4b	
(1)	4c	5

N=68

Mean = 1.35 Median = 1.00 SD = .946

IR = .820 .75IR = .615 Skewness = 1.733

(kurtosis -3) = 2.837

[a]Values greater than 2.5 were recoded as missing for analytic purposes.

Figure D-3

Stem-and Leaf Display, Average Number of Parent Education Sessions Per Month

(a for leaf 0, 1; b for leaf 2, 3; c for leaf 4, 5; d for leaf 6, 7; e is leaf 8, 9)

N		
(7)	0b	2233333
(7)	0c	4444555
(12)	0d	666666667777
(5)	0e	88889
(25)	1a	0000000000000000000000011
(5)	1b	22333
(3)	1c	455
(2)	1d	67

N=66

Mean = .85   Median = 1.00   SD = .355

IR = .430   .75IR = .322   Skewness = .103

(kurtosis -3) = -.392

Figure D-4

Stem-and Leaf Display,
Number of I/T Sessions[a]

N		
(13)	1	0000000000005
(4)	2	0000
(12)	3	000000000000
(6)	4	000005
(2)	5	00
(5)	6	00000
(6)	7	000000
(3)	8	000
(3)	9	005
(2)	10	05
(3)	11	000
(2)	12	00
(2)	13	00
(1)	14	0
(5)	15	00005
(0)	16	
(1)	17	0
(2)	18	05
(4)	19	0055
(1)	20	0
(0)	21	
(1)	22	0

N=78

Mean = 7.4 Median = 6 SD = 5.96

IR = 8.0 .75IR = 6.0 Skewness = 0.81

(kurtosis -3) = -0.49

[a]Under certain conditions, counts of 0.5 were permitted.

Figure D-5

Stem-and Leaf Display,
Number of Parent Education Sessions[a]

N		
(17)	1	00000000000000000
(7)	2	0000000
(4)	3	0000
(7)	4	0000005
(3)	5	000
(3)	6	000
(1)	7	0
(5)	8	00005
(3)	9	000
(3)	10	000
(4)	11	0000
(0)	12	
(2)	13	00
(1)	14	0
(0)	15	
(1)	16	0
(1)	17	0
(3)	18	005

N=65

Mean = 5.9 Median = 4 SD = 5.04
IR = 8.0 .75IR = 6.0 Skewness = 0.97
(kurtosis -3) = 0.05

[a]Under certain conditions, counts of 0.5 were permitted.

Figure D-6

Stem-and Leaf Display,
Number of Center Sessions, I/T and Parent Averaged[a]

N		
(19)	1	0000000000000000025
(9)	2	000000055
(10)	3	0000000055
(6)	4	000005
(7)	5	0000255
(6)	6	000555
(4)	7	0002
(3)	8	005
(2)	9	00
(3)	10	255
(4)	11	0005
(1)	12	0
(2)	13	05
(2)	14	00
(4)	15	0025
(0)	16	
(1)	17	0
(4)	18	0055
(0)	19	
(1)	20	0

N=88

Mean = 6.3 Median = 4.75 SD = 5.29

IR = 7.0 .75IR = 5.25 Skewness = 0.98

(kurtosis -3) = -0.10

[a]If a family was missing data on only one of the two total participation measures averaged here, the value for the other was permitted to stand.

Figure D-7

Stem-and Leaf Display,
Natural Log of Number of Center Sessions

(a for leaf 0, 1; b for leaf 2, 3; c for leaf 4, 5;
d for leaf 6, 7; e for leaf 8, 9)

N		
(17)	0a	00000000000000000
(1)	0b	2
(1)	0c	4
(7)	0d	7777777
(2)	0e	99
(8)	1a	11111111
(2)	1b	33
(6)	1c	444445
(7)	1d	6666777
(9)	1e	888999999
(4)	2a	0111
(3)	2b	223
(7)	2c	4444445
(8)	2d	66667777
(5)	2e	89999
(1)	3a	0

N=88

Mean = 1.43 Median = 1.56 SD = 0.96

IR = 1.50 .75IR = 1.13 Skewness = -0.15

(kurtosis -3) = -1.18

13

D.2 Measuring Family Needs and Strengths

Program participation as well as developmental scores may be related to various family needs or strengths. Much of the data available concerning needs and strengths comes from CFRP staff reports, or from interviews conducted only with CFRP mothers; the modeling of developmental scores (Appendix C) could not take these measures into account, since no control/comparison group data are available. In attempting to understand relationships between participation in CFRP and developmental outcomes, however, they must be considered.

A total of nine needs or strengths variables were identified or constructed. All are family-level measures. Some derive from parents' self-reports, others from staff reports; some originate in data collected during the fall 1978 (baseline) field effort, others in the spring 1979 data collection.

Two measures assess needs in a general way. One (HASSLED) is from the baseline parent interview, where a series of items asked how frequently the respondent felt "hassled" or bothered by various types of people (neighbors, doctor, friends, family, and relatives): HASSLED combines these items, resulting in a single, self-reported measure of the frequency with which one feels "hassled" generally (the response scale for each <u>item</u> ranged, in four steps, from "almost never" to "weekly," but the HASSLED scale retains only the <u>direction</u> of the item responses, from lesser to greater frequency). The second measure may be considered to be a need index (NEEDS), and is based on staff reports collected on three instruments, one in the fall 1978 and two in the spring 1979 data collections. All items were of a checklist form, each item asking about specific needs (in areas of housing, health, income, family management, etc.); subsequent analyses of these items failed to produce a set of relatively independent, content-based measures of need, so they were combined into one index. The NEEDS scale has no inherent metric; a higher score indicates greater need. Figures D-8 and D-9 give stem-and-leaf displays for HASSLED and NEEDS.

Three measures assess family strengths with respect to CFRP participation. The first (CFRP SUPPORT) combines items from the fall 1978 and spring 1979 parent interviews. Two items in each interview asked about the degree of support for CFRP involvement from family and friends; these were combined into one measure, with a higher score indicating greater support. The second measure (RESOURCES) is an assessment of personal or social resources available to the family, and comes from the spring 1979 CFRP staff interviews. It includes items considering the frequency of the mother's social contacts outside of CFRP, her ties with an extended family, and her awareness of social services available in the community. The third measure of strengths (ENTHUSIASM) addresses enthusiasm for the program, and also derives from items in the spring 1979 staff interviews. Staff were asked to rate mothers' degree of interest or enthusiasm for various CFRP activities; separate constructs by types of activities were impossible to distinguish, so one measure was created combining these items. During recent (spring 1980) interviews with staff from all eleven CFRPs, the concept of individual motivation--although phrased differently by different staff--was mentioned universally as the single most important ingredient of success in working with families; ENTHUSIASM seems to be the only measure available as a motivation surrogate. These variables are described in Figures D-10 through D-12.

The remaining four variables are ambiguous in their valence: it is not clear whether a high score is more desirable than a lower score. The first three all have to do with the

Figure D-8

Stem-and Leaf Display, HASSLED[a]

(stem units are tenths)

N		
(18)	0	000000000000000000
	1	
(27)	2	000000000000000000000000000
	3	
(26)	4	00000000000000000000000000
(2)	5	00
(18)	6	000000000000000000
(3)	7	555
(24)	8	000000000000000000000000
	9	
(25)	10	0000000000000000000000000
N=143		

Mean = .52 Median = .50 SD = .335

IR = .600 .75IR = .450 Skewness = -.004

(kurtosis -3) = -1.260

---

[a]A higher score indicates greater reported frequency of feeling "hassled" by others.

Figure D-9

Stem-and Leaf Display, NEEDS[a]

(a for leaf 0, 1; b for leaf 2, 3; c for leaf 4, 5; d for leaf 6, 7; e for leaf 8, 9)

N		
(3)	0c	455
(3)	0d	667
(7)	0e	8888999
(14)	1a	00000000000011
(18)	1b	222222222333333333
(13)	1c	4444455555555
(29)	1d	66666666666666777777777777777
(12	1e	888888899999
(13)	2a	0000111111111
(9)	2b	222233333
(2)	12c	44
(2)	2d	77
(1)	2e	9
N=126		

Mean = 1.55 Median = 1.60 SD = .496

IR = .675 .75IR = .506 Skewness = -.038

(kurtosis -3) = -.316

[a] A higher score indicates greater need.

Figure D-10

Stem-and Leaf Display, CFRP SUPPORT[a]

(a for leaf 0,1; b for leaf 2,3; c for leaf 4,5; d for leaf 6,7; e for leaf 8,9)

N		
(3)	1a	111
(16)	1b	2222222222222222
(8)	1c	55555555
(6)	1d	667777
(10)	1e	8888888899
(4)	2a	0000
(22)	2b	2333333333333333333333
(5)	2c	44444
(11)	2d	66666666777
(14)	2c	88999999999999
(7)	3a	0111111
(4)	3b	2222
(16)	3c	4444444444444445
N=126		

Mean = 2.31 Median = 2.32 SD = .741

IR = 1.150 .75IR = .862 Skewness = -.077

(kurtosis -3) = -1.142

[a]A higher score indicates greater support.

Figure D-11

Stem-and Leaf Display, RESOURCES[a]

(stem units are tenths)

N		
(6)	2	555555
(0)	3	
(0)	4	
(16)	5	0000000000000000
(3)	6	777
(30)	7	555555555555555555555555555555
(0)	8	
(0)	9	
(34)	10	0000000000000000000000000000000000
(0)	11	
(24)	12	555555555555555555555555
(5)	13	33333
(0)	14	
(7)	15	0000000
(1)	16	7

N=126

Mean = .93 Median = 1.00 SD = .325

IR = .500 .75IR = .375 Skewness = -.109

(kurtosis -3) = -.586

[a]A higher score implies greater personal resources available.

Figure D-12

Stem-and Leaf Display, ENTHUSIASM[a]

(a for leaf unit 0,1; b for leaf unit 2,3; c for leaf unit 4,5; d for leaf unit 6,7; e is leaf unit 8,9)

N		
(7)	1a	0011111
(9)	1b	222666666
(4)	1c	4445
(11)	1d	66667777777
(11)	1e	88888999999
(12)	2a	000000111111
(18)	2b	222222223333333333
(9)	2c	444445555
(14)	2d	66666666777777
(12)	2e	888899999999
(6)	3a	000011
(6)	3b	222333
(11)	3c	44444555555

N=130

Mean = 2.29 Median = 2.30 SD = .682

IR = 1.000 .75IR = .750 Skewness = -.012

(kurtosis -3) = -.813

[a]A higher score indicates greater enthusiasm for or interest in CFRP.

degree of help (from various sources) the mother has in dealing with the routine tasks of child care; these come from the fall 1978 parent interview. The first, called SPOUSE HELPS, assesses the extent to which the mother's spouse (or live-in male partner) helps. The second assesses help from other relatives (RELATIVE HELPS). The third combined responses concerning older children, neighbors, and babysitters (OTHERS HELP). It is not clear, prima facie, whether help with more tasks in any category indicates a positive resource--the availability of someone to rely upon for support--or a negative condition--that the mother gets additional help in whatever form because she cannot or does not sufficiently manage these routine tasks for herself. There was a relatively high incidence of missing data on the series of HELP items; the constructs are summarized in Figures D-13 through D-15.

The fourth measure comes from the spring 1979 staff interviews, and assesses the extent to which child development content was emphasized in working with the family during the initial (fall 1978 to spring 1979) phase of CFRP involvement (CD EMPHASIS). Again, a high score could represent either a positive or negative phenomenon: a family's needs may have been met sufficiently to devote relatively more time to child development activities, or the family's needs in the area of child development (nutrition and basic physical care) may have been so severe as to warrant relatively extensive attention. This variable has only three response categories; 58 families scored zero, minimum emphasis; 55 families scored one; 22 families scored two, maximum emphasis.

Table D-2 shows correlations among the nine needs/strengths variables. While these variables are not independent, quite clearly they represent a multidimensional set of measures. No further reduction of this set of measures seems

Figure D-13

Stem-and Leaf Display, SPOUSE HELPS[a]

(stem units are tenths)

N		
(8)	1	22223333
(9)	2	555555555
(13)	3	7777778888888
(0)	4	
(11)	5	00000000000
(11)	6	22222233333
(7)	7	5555555
(7)	8	7778888
(0)	9	
(2)	10	00

N=68

Mean = .50 Median = .50 SD = .244

IR = .375 .75IR = .281 Skewness = -.195

(kurtosis -3) = -.871

[a]A higher score indicates more help with routine child care tasks.

Figure D-14

Stem-and Leaf Display, RELATIVE HELPS[a]

(stem units are tenths)

N		
(10)	1	2222233333
(11)	2	55555555555
(12)	3	777777888888
(0)	4	
(6)	5	000000
(10)	6	2222233333
(18)	7	555555555555555555
(12)	8	777777888888
(0)	9	
(10)	10	0000000000

N=89

Mean = .58   Median = .62   SD = .285

IR = .375   .75IR = .281   Skewness = -.169

(kurtosis -3) = -1.268

[a]A higher score indicates more help with routine child care tasks.

Figure D-15

Stem-and Leaf Display, OTHERS HELP[a]

(a for leaf 0,1; b for leaf 2,3; c for leaf 4,5; d for leaf 6,7; e for leaf 8,9)

N		
(12)	-2a	111111111111
(8)	-1e	99999999
(3)	-1d	777
(16)	-1c	4444444444444444
(6)	-1f	333322
(9)	-1a	111000000
(1)	-0e	8
(6)	-0d	777776
(4)	-0c	5555
(2)	-0b	32
(1)	-0a	1

N=68

Mean = -1.58 Median = -1.38 SD = .557

IR = .965 .75IR = .724 Skewness = -.279

(kurtosis -3) = -.830

[a]The natural logarithm of this variable is used throughout. A higher score indicates more help with routine child care tasks.

Table D-2

Correlations Among Needs/Strengths Variables[a]

	HASSLED	NEEDS	CFRP SUPPORT	RE-SOURCES	ENTHU-SIASM	SPOUSE HELPS	RELATIVE HELPS	OTHERS HELP
NEEDS	-.13							
CFRP SUPPORT	.09	.26						
RESOURCES	.21	.10	.12					
ENTHUSIASM	.01	.32	.21	.16				
SPOUSE HELPS	-.04	-.17	.05	.04	-.14			
RELATIVE HELPS	.11	.03	-.14	-.12	.02	-.06		
OTHERS HELP[b]	.13	.31	-.03	.09	.29	.06	.22	
CD EMPHASIS	-.10	-.01	-.07	-.00	-.14	.01	.05	.02

[a] Product-moment correlations.

[b] The natural logarithm of OTHERS HELP was used.

warranted, although CFRP SUPPORT, NEEDS, ENTHUSIASM, and OTHERS HELP seem to be moderately (positively) intercorrelated.

Table D-3 lists correlations between the needs/strengths set of variables and the five basic participation measures constructed. Clearly, any understanding of participation in CFRP--a topic in the process/treatment study, to be addressed in a later report--must consider family needs and strengths. There are suggestions, and they cannot be followed-up here sufficiently, that: the fewer personal resources available the more home visits received; support for CFRP involvement may increase rates of participation in center sessions as well as the number of center sessions attended, but not whether center activities are avoided altogether; center session attendance <u>in part</u> may substitute for assistance from others in caring for children; and staff emphasis on child development in working with mothers may be a reaction to relatively infrequent attendance at I/T sessions.

Table D-3

Correlations Between Needs/Strengths and Participation[a]

	Home Visit Rate	I/T Session Rate	Parent Ed. Session Rate	Total Center Participation[b]	Center Participation Dummy
HASSLED	-.01	.09	-.25	.01	.04
CFRP SUPPORT	.07	.20	.16	.29	.08
RESOURCES	-.22	.11	.14	.00	-.10
ENTHUSIASM	.17	.14	.14	.29	.36
SPOUSE HELPS	.02	.19	-.23	-.10	-.20
RELATIVE HELPS	-.18	-.30	-.23	.12	-.21
OTHERS HELP[b]	-.04	-.39	.05	-.07	-.08
CD EMPHASIS	.08	-.36	.11	.02	.02

[a]Product-moment correlations.

[b]The natural logarithms of total center participation and OTHERS HELP were used.

14

D.3 <u>Program Participation and Developmental Benefits</u>

For two major reasons, the exploration of relationships between participation in CFRP and developmental benefits must be done within the CFRP group only. First, participation measures are available only in the CFRP group. Given that the control/comparison group families do not participate in CFRP, it might seem reasonable at first glance to impute zero values for all participation measures and to include them in all analyses. Given the evaluation design, and indeed the CFRP demonstration intent, this would be wrong. CFRP is not necessarily intended to be the <u>only</u> developmentally beneficial program available to eligible parents in the demonstration communities; day care, developmental screening, and health and nutrition services are available to control/comparison families. The analyses reported in Appendix C addressed the question of whether CFRP involvement produces developmental gains (as measured by the Bayley Scales) relative to <u>whatever</u> services are available to and used by other families in the community. To impute zero values for participation to control/comparison families without otherwise "adjusting" for any relevant services they may be receiving would attenuate estimates of relationships of participation with developmental gains and obscure anything we might otherwise learn about CFRP benefits to children.

The second reason for limiting these explorations to CFRP families is simply that we have no comparable needs or strengths measures on control/comparison families. Inasmuch as program participation is a complex phenomenon, requiring some way to include measures of needs and/or strengths in understanding the relationship between participation and developmental benefits, the control/comparison families will be excluded <u>de facto</u>. The exploratory analyses reported here, therefore, are restricted to CFRP families.

Finally, except for analyses of within-Salem Mental Score value-added estimates, participation analyses were done globally, across all sites. The total sample available is 151, and with 95 to 146 valid scores on home visits and developmental scores, leaving very small samples for simple analyses within sites. When center participation measures are considered, the total sample is less than 70; with certain interactions, it drops to about 30. Analyses employing these variables cannot be done usefully within sites.

The dependent variables used are residuals and child-level value-added estimates from the developmental models estimated and reported in Appendix C. There are five variables: global ANCOVA Mental and Physical Score residuals, global Mental and Physical Score value-added growth model estimates, and within-Salem Mental Score value-added growth model estimates. Table D-4 gives descriptive statistics for these measures; Table D-5 contains their correlation matrix. Clearly, the two Physical Score measures cannot be expected to provide results differing by the growth model adopted (ANCOVA or value-added): their correlation is .87. The within-Salem Mental Scores, too, are highly correlated with their global counterparts: correlations are .82 with the global value-added estimates and .91 with the global ANCOVA residuals. The two global Mental Score measures are only moderately correlated (.68), and therefore they may yield different results according to the global growth model adopted. The global Physical Score measures are correlated only minimally with the global Mental Score measure, leaving the possibility that participation may be related to Physical and Mental Score measures in different ways.

Participation Models

There is no theory of just how program participation *ought* to be related to developmental gains, and a fair

Table D-4

Developmental Outcomes for Participation Exploration

Variable (N)	Median	Mean	SD	Skewness	(Kurtosis -3)
Mental Score Value-Added, or VADDGM[a] (95)	-0.23	-0.49	5.64	-0.09	-0.87
ANCOVA Mental Score Residual, or ANCOVAGM[a] (102)	0.16	0.37	4.52	0.08	-0.58
Within-Salem Mental Score Value-Added, or VADDED.M[a] (17)	3.03	2.18	4.51	-0.70	-0.12
Physical Score Value-Added, or VADDGP[b] (110)	-0.015	-0.019	0.197	0.049	-0.889
ANCOVA Physical Score Residual, or ANCOVAGP (102)	-0.030	-0.018	0.196	-0.025	-0.409

[a]The scale is that of the Mental Score analyses reported in Section 3.2--that is, points on the Bayley Mental Score.

[b]The scale is that of the Physical Score analyses reported in Section 3.3--that is, points on the Bayley Physical Score. (One-tenth of a point is one test item.)

Table D-5

Correlations Among Developmental Outcomes[a]

	VADDGM	ANCOVAGM	VADDED.M[b]	VADDGP
ANCOVAGM	.68			
VADDED.M[b]	.82	.91		
VADDGP	.26	.20	.43	
ANCOVAGP	.29	.19	.52	.87

[a]Product-moment correlations.

[b]Correlations within Salem only.

14

number of plausible arguments can be made a priori. The simplest model conceptually is the easiest to investigate empirically, but it suffers from obvious flaws. This model posits greater participation leading to greater developmental gains ("more is better"). This is not totally implausible: if some proportion of "activity time" (be it home visits or center sessions) is devoted to working with parents and child together (say, to demonstrate age-appropriate, developmentally stimulating activities), and if parents' effectiveness in learning and using these activities is a simple function of how many (not how often) demonstrations they have been included in, then a "more is better" hypothesis would be justified. A more reasonable alternative, however, suggests that the relative frequency of participation--the rate at which families/children participate--is important. Relatively regular interaction may be required for the intended benefit to "take." The simplest version of this model is a "more is better" variant: more frequent participation rather than simply more participation should lead to greater developmental benefits.

Both versions of this "more is better" model are deficient, however, in that they fail to account for possible relationships between types of activities in CFRP. There are three such types of possible developmental importance--home visits (HV), parent education (PED) sessions, and infant-toddler (I/T) sessions. (The distinction between PED and I/T sessions is somewhat vague, empirically. The idea was to count sessions involving children separately from those involving parents only. The distinction in practice is not that clear, and program staff often had difficulty in recording any given center session attendance as only one of these two types. Nevertheless, some distinction is evident in the data.) The two versions of a "more is better" model simply suggest that more or more frequent participation in any of these activities would lead to greater developmental gains. The CFRP philosophy suggests that home visits and center sessions ought to be integrated and/or complementary in their focus and in their intended effect. That is, the effects of participating in any given type of activity ought not to be independent of the effects due to participation in other activities. The model presumed (at least tacitly) is an interactive one: the benefits expected when a family participates more (or more frequently) in multiple kinds of CFRP activities should exceed the benefits expected from each kind of activity alone. These might be called "more or more frequently in combination is better" models.

Thus far, four (not necessarily mutually contradictory or exclusive) simple models of the participation-developmental impact relationship have been identified. In ascending order of complexity, they are:

- Increased participation increases developmental gains (or "more is better").
- Increased rates of participation increase developmental gains (or "more frequently is better").
- Increased participation increases developmental gains, but increased participation in both center and home visit activities increases developmental gains even more (or "more in combination is better").
- Increased rates of participation increase developmental gains, but increased participation rates in both center and home visit activities increase developmental gains even more (or "more frequently in combination is better").

These models are "simple" only in the sense that they fail to take any account of family needs or strengths. Since needs and/or strengths quite possibly affect CFRP participation and developmental gains made through such participation, needs and strengths ought to be considered when modeling any relationship between CFRP participation and developmental outcomes. Again, we have no theory to specify how needs and strengths should affect the relationship (or, for that matter, just which needs and strengths to measure). Three basic ways of including needs and strengths can be identified. First, it is clear that CFRP families vary in the type and degree of needs that they have, and that participation (whether total amounts, or rates of participation) may vary according to family needs. It is quite possible, therefore, that a standardization of participation measures relative to need might be useful: the model here is that increased participation or participation rate <u>per unit of need</u> leads to greater developmental gains. In the CFRP evaluation data, only one general extent-of-needs variable could be distinguished (NEEDS), and a set of participation variables (total amounts and rates) were standardized relative to it. The resulting constructs were so highly correlated with their simpler, unstandardized counterparts that the standardized versions cannot be judged to contain any additional information. (In all but one instance, rank order correlations exceeded .85.) Needs-standardized measures were dropped from consideration.

A second conception of how needs and strengths affect participation and developmental outcomes is that of the simple linear model: needs and/or strengths covary with both participation and developmental gains, are causally prir to both, and must be "controlled" or "adjusted for" in examining the relationship between participation and developmental impact. The analytic strategy suggested here is that of regression modeling of outcomes on needs and strengths first, adding participation variables later (with appropriate needs and/or strengths "controlled").

The third way in which needs or strengths variables can fit into a model of the participation/development relationship is interactive. The participation _effect_ itself may vary with family needs and/or strengths. For instance, increased participation may lead to _greater_ developmental gains for families with fewer needs or more strengths than for families with more needs or fewer strengths. The appropriate independent variables under this model are products (interactions) of participation measures with needs and strengths variables.

All told, therefore, there are _twelve sets of models_ potentially of some use in exploring the relationship between participation and developmental gains: four "simple" models, as listed earlier; four corresponding models in which needs and strengths must be considered as covariates of developmental outcomes; and four corresponding models in which participation interactions with needs and strengths are explored. Without a strong theory from which to argue the inappropriateness of any model, all must be considered in these exploratory analyses. Furthermore, with sample sizes as small as those available, and given the problems of collinearity often faced in trying to estimate (and compare) main effects and interactions simultaneously in regression models, we cannot possibly hope to construct "critical tests" of these alternative models here. All we can do is look for indications that participation may effect developmental outcomes in a manner consistent with the predictions of any of these twelve models. Without a strong theory to back any exclusions _a priori_, all must be considered.

Results

The "simple" models are easiest to explore, since they do not consider needs and strengths measures. The correlations between developmental outcome scores and total participation, participation rates, and various participation interactions are given in Table D-6. (Examination of

Table D-6

Correlations Between Developmental Scores and Various Participation Measures[a]

	VADDGM	ANCOVAGM	VADDED.M[b]	VADDGP	ANCOVAGP
Home Visit Rate	-.02	.06	.02	.11	.17
I/T Session Rate	.10	.39	.19	.32	.27
Parent Ed. Session Rate	-.05	.05	.20	-.07	-.16
Center Participation Dummy[c]	.06	-.00	.25	.20	.13
Total Center Participation[d]	.04	.14	.48	.02	.15
Interactions					
Home Visit Rate by Center Participation Dummy	.03	.04	.25	.13	.16
Home Visit Rate by Total Center Participation	-.05	.04	.16	.03	.16
Home Visit Rate by I/T Session Rate	-.02	.41	.18	.13	.24
Home Visit Rate by Parent Ed. Session Rate	-.05	.05	.10	-.17	.01
I/T Session Rate by Total Center Participation	.02	.29	.15	.31	.26
Parent Ed. Session Rate by Total Center Participation	.01	-.06	.17	.02	-.12

[a] Sample sizes vary, according to pairs of variables. In general, for global outcomes with the home visit rate and center participation dummy variables, N is about 95; within Salem, about 25. For global outcomes with I/T session rate, parent education session rate, and total center participation, the sample size is about 60; within Salem, approximately 15. Sample sizes for correlations involving interactions may be somewhat smaller. (For sample sizes of 25 and 15, correlations of .22 and .29 are significant in one-tailed tests at the .15 level; for Ns of 95 and 60, with one-tailed tests at the .05 level, the appropriate correlations are .17 and .21.)

[b] These are within-Salem only.

[c] This is a dichotomous variable.

[d] The natural logarithm of this variable was used.

scatterplots revealed few outliers, which were screened before computing these correlations, and no evidence of nonlinearity in these bivariate relationships.) Although there are no indicators of significant correlations between participation and Mental Score outcomes under the global value-added model, there are some positive signs for each of the other four outcome possibilities.

The I/T session attendance rate is positively correlated with both versions of the Physical Score gains, and with the ANCOVA model Mental Score gains. Total center participation and perhaps the center participation dummy are associated with value-added model Mental Score gains in Salem. Among the participation interaction variables, two stand out: the interaction between home visit and I/T session rates is correlated fairly highly with the ANCOVA Mental and Physical score residuals; and the interaction between I/T session attendance rate and total center participation--in essence, greater frequency of attendance over longer periods of time--is positively correlated with ANCOVA Mental Score residuals and with both versions of Physical Score gains.

If regression models are estimated using the entire set of participation measures as potential covariates, the resulting models are quite parsimonious. Since the I/T session rate is highly correlated with the home visit rate-by-I/T session rate interaction (.75) and with the I/T session rate-by-total center participation (.92), models for the ANCOVA residuals cannot be estimated usefully with more than one of these three variables in the equation. Alternative models for each outcome can be estimated, one with the I/T session rate and another with the appropriate interaction term. From the regressions estimated, predictions of typical gains given specific values of the participation variable can be made. Predicted gains for simple models (with I/T participation rate as the only predictor) are shown in Table D-7.

Table D-7

Estimated Gains from I/T Sessions per Month

I/T Session Rate[c]	ANCOVAGM[a]	ANCOVAGP[b]	VADDGP[b]
0.6	-1.36	-.070	-.081
0.8	-0.64	-.048	-.055
1.0	0.07	-.027	-.030
1.2	0.79	-.005	-.004
1.4	1.51	.016	.021
1.6	2.23	.037	.047
1.8	2.95	.059	.073
2.0	3.66	.080	.098

[a]These are points (or number of items) on the Bayley Mental Score.

[b]These are tenths of a point (or item) on the Bayley Physical Score.

[c]These are average number of I/T sessions per month. The sample mean is about 1.1 I/T sessions per month.

If these models have any validity, it seems that the participation rate in I/T sessions must reach about three sessions every two months to produce positive gains on the Physical Score, and should be at (or greater than) two sessions per month for these gains to be appreciable--that is, to approach an average increase of one full item on the 10-item scale. For gains on Mental Scores based on the ANCOVA modeling, once per month seems to produce a very small positive gain, while the average participation rate should exceed three sessions every two months for appreciable Mental Score gains. (In the evaluation sample, an average CFRP gain of about <u>two</u> points at 18 months of age--approximately the sample mean age--would put CFRP children at the national norm for that age. Of course, as they get older, gains relative to a comparable non-CFRP population would have to be greater to keep the children at the national norm.)

From the within-Salem value-added Mental Score regression, attendance at roughly 12 center sessions to the date of Bayley testing (combining both I/T and parent education sessions) predicts a typical Mental Score gain of about three points. Most of the children were tested after about one year in the program, so this amounts to only one center session per month, in the within-Salem model. (This relationship was estimated on a sample of only 22 children, however, and should not be given excessive weight.)

The introduction of needs and strengths variables has little effect on the relationships estimated in the "simple" models above. Correlations between needs/strengths and developmental scores are given in Table D-8. Three general points stand out:

- NEEDS, RESOURCES, and ENTHUSIASM are all positively correlated with Mental Score outcomes, but not with Physical Score outcomes.

- CD EMPHASIS is correlated negatively with Mental Score gains, suggesting (but only suggesting) that staff emphasis on child development activities may reflect perceived need in this area rather than strength in other areas and concomitant opportunity to move beyond more basic family concerns.

- RELATIVE HELPS is negatively correlated with each of the developmental outcomes, suggesting that what was measured here is more of a developmental liability than an asset.

In regressions on Physical Score gains, with needs and strengths taking precedence over participation measures initially, the relationships between participation and Physical Score are unaffected. In the within-Salem analysis, the NEEDS variable enters first, but the total center participation measure is still strongly and positively associated with value-added Mental Score gains (the partial correlation, with NEEDS controlled, is about .50, up from .48). In the ANCOVA-based Mental Score gains, RESOURCES and

Table D-8

Correlations Between Needs/Strengths Measures and Developmental Gains[a]

	VADDGM	ANCOVAGM	VADDED.M[b]	VADDGP	ANCOVAGP
HASSLED	.09	-.01	-.11	.01	.00
NEEDS	.18	.07	.33	.06	.01
CFRP SUPPORT	.02	.10	.20	-.02	-.09
RESOURCES	.12	.24	.09	.13	.06
ENTHUSIASM	.08	.18	.25	.09	.04
SPOUSE HELPS[c]	-.03	.02	-.31	.03	-.14
RELATIVE HELPS[c]	-.26	-.19	-.38	-.22	-.24
OTHERS HELP[c,d]	-.07	.03	-.16	.06	-.06
CD EMPHASIS	-.21	-.23	-.08	-.12	-.03

[a]The needs and strengths variables are described in Section D.2.

[b]Within-Salem only.

[c]These variables had high proportions of missing data and were excluded from regression analyses.

[d]The natural logarithm of OTHERS HELP was used.

CD EMPHASIS both enter the regression, but I/T session attendance rate is still strongly related to the outcome measure (the partial correlation with RESOURCES and CD EMPHASIS controlled is .33, down from .39).

The final set of models to be explored focus on interactions between participation and needs/strengths. With six needs and strengths variables (the HELP series was omitted due to large proportions of missing data) and five "simple" participation measures, there are 30 "simple" interactions to explore. From among this set, four bivariate relationships are suggestive, each for a different outcome variable with the interaction between I/T session rate and RESOURCES: with VADDGM, the correlation is .34; with ANCOVAGM, .41; with VADDGP, .43; and with ANCOVAGP, .34. These correlations, and the regression models estimated as a result of exploratory analyses, suggest that the impact of rate of attendance at I/T sessions varies with the personal or social resources available to the mother, so that developmental gains due to increased rate of I/T session attendance would be greater among children whose mothers have relatively more resources available to them. The effect of this interaction is shown in Tables D-9 and D-10.

Table D-9

Estimated Physical Score Value-Added, by RESOURCES and I/T Session Rate[a]

Average Number of I/T Sessions per Month	RESOURCES Low[b]	Medium[b]	High[b]
0.50	-.125	-.095	-.065
0.75	-.097	-.052	-.006
1.00	-.069	-.009	.052
1.25	-.041	.034	.110
1.50	-.012	.077	.168

[a]Table entries are tenths of a point (or item) on the Bayley Physical Score, a 10-item scale.

[b]Low RESOURCES are one standard deviation below the sample mean; medium RESOURCES are at the sample mean; high RESOURCES are one standard deviation above the mean.

Table D-10

Estimated Mental Score Gains from the ANCOVA Model, by RESOURCES and I/T Session Rates[a]

Average Number of I/T Sessions per Month	RESOURCES Low[b]	Medium[b]	High[b]
0.50	-1.93	-1.27	-0.62
0.75	-1.32	-0.33	0.65
1.00	-0.71	0.60	1.91
1.25	-0.10	1.54	3.18
1.50	0.51	2.48	4.44

[a]Table entries are points (or item) on the Bayley Mental Score.

[b]Low RESOURCES are one standard deviation below the sample mean; medium RESOURCES are at the sample mean; high RESOURCES are one standard deviation above the mean.

In reading Tables D-9 and D-10, it is important to remember that the average rate of participation in I/T sessions is just over once per month. The distribution of sample families by RESOURCES and I/T participation rates is centered around the values in the middle of these tables; as values change in any direction, the predicted "gains" are more unstable, less reliable, or less accurate. Nevertheless, if these models have some validity, it is clear that participation rates in I/T sessions can influence developmental scores appreciably--especially if the home visiting activities are successful in increasing the mother's awareness of social services available in the community and in creating and/or strengthening the family's ties to supportive social networks (these kinds of items make up the RESOURCES construct). An average gain of even one-third to one-half of a point on the Bayley Physical Score is impressive (the scale used consists of ten items altogether); as mentioned earlier, an average gain of about two points on the Mental Score, at 18 months of age, would put the CFRP sample at the national norm.

A final set of explorations was done, checking third-order interactions. These were home visiting rate by I/T session rate by each of three needs/strengths variables: NEEDS, CFRP SUPPORT, and RESOURCES. From these analyses, only the global value-added Mental Score outcomes are related sensibly to the interaction with RESOURCES. (The sensibility of exploring interactions of this order is debatable, especially without a theory to suggest that they should take precedence over simpler expressions of needs, strengths, and participation. Nevertheless, these particular interactions do have a plausible backing in the CFRP application.) This suggests that greater RESOURCES, more frequent home visits, and more frequent I/T session attendance each mutually reinforces or "levers" the impact of the others.

Summary

Despite inability to detect group differences in Bayley developmental scores (CFRP versus control/comparison children) after 12 to 18 months of enrollment, it appears that certain aspects of CFRP participation could be expected to produce demonstrable, important developmental gains under appropriate conditions. Increased participation in center sessions, particularly I/T sessions, evidently can enhance children's development. Depending on the participation model chosen as most appropriate, center session participation at rates that are somewhat higher than those currently found--at least in the evaluation sample--and more widespread attendance at center sessions (approximately half of the evaluation families had never attended a center session as of the Bayley testing) conceivably could produce demonstrable impacts.

Potential interactions with RESOURCES and with home visiting rates further attest to the potential in the CFRP concept. From limited empirical evidence, it seems that an integrated, family-focused approach which employs a mixture of home visiting and center sessions could produce important developmental impacts among children in CFRP-eligible families at ages as young as 18 months. The sample of families in the CFRP evaluation, however, have so far not participated with anywhere near the frequency required to offer any hope of detecting such impacts.

169