DOCUMENT RESUME

ED 129 416 PS 008 809

AUTHOR Datta, Lois-ellin; And Others
TITLE The Effects of the Head Start Classroom Experience on Some Aspects of Child Development: A Summary Report of National Evaluations, 1966-1969.
INSTITUTION Office of Human Development (DHEW), Washington, D.C.
REPORT NO DHEW-OHD-76-30088
PUB DATE [76]
NOTE 92p.

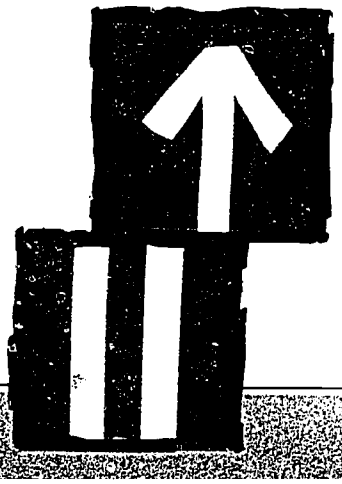
EDRS PRICE MF-$0.83 HC-$4.67 Plus Postage.
DESCRIPTORS Cognitive Development; *Comparative Analysis; *Disadvantaged Youth; *Early Childhood Education; Evaluation Methods; Family Characteristics; Family Involvement; Federal Programs; *Intervention; Measurement Techniques; Motivation; *Preschool Programs; Program Effectiveness; *Program Evaluation; Readiness (Mental); Research; Social Development; Statistical Analysis; Student Characteristics; Teacher Characteristics
IDENTIFIERS *Project Head Start

ABSTRACT

This report summarizes the findings of previous technical reports on the immediate changes in child development associated with Head Start and presents a statement of the extent to which these changes have been enhanced or attenuated by different program experiences for different children. Background information presented includes an overview of the findings; a history of Project Head Start; and an overview of Head Start Evaluations (descriptive studies, on-site research and evaluation studies, national impact studies, national interaction model studies). A section devoted to design and measurement examines methodological issues, design issues, data collection, measures, and statistical analytic models. A section on findings covers questions concerned with characteristics of Head Start children and families, effectiveness of Head Start programs and teacher approaches, and children's performance gains. A discussion section examines factors influencing developmental changes and implications for early childhood intervention. Appendices are (1) Measure Selection and Psychometric Characteristics and (2) Selection of Variables for Analysis. (JH)

THE EFFECTS OF THE HEAD START CLASSROOM EXPERIENCE ON SOME ASPECTS OF CHILD DEVELOPMENT:

A Summary Report of National Evaluations 1966-1969



THE EFFECTS OF THE HEAD START CLASSROOM EXPERIENCE ON SOME ASPECTS OF CHILD DEVELOPMENT: A SUMMARY REPORT OF NATIONAL EVALUATIONS 1966-1969

Lois-ellin Datta
Carol McHale
Sandra Mitchell

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Office of Human Development
Office of Child Development

DHEW Publication No. (OHD) 76-30088

TABLE OF CONTENTS

BACKGROUND 1

Introduction to the Studies 1

Overview of the Findings 3

History of Project Head Start 5

Overview of Head Start Evaluations 10

1. Descriptive Studies 11

2. One-Site Research and Evaluation Studies 11

3. National Impact Studies 13

4. National Interaction Model Studies 13

DESIGN AND MEASUREMENT 16

Methodological Issues 16

Design Issues 17

Data Collection 19

Measures 22

Statistical Analytic Models 26

FINDINGS 29

What Were the Head Start Children and Families Like? 29

Were Differences Among Children's Prescores Related to Background Characteristics of the Children and Families 30

What Were the Head Start Programs Like? 33

Did Children Gain From Their Head Start Experience? 38

Which Child and Family Characteristics Are Associated with the Greatest Gains in Performance? 44

Which Teacher and Program Characteristics Are Associated with the Greatest Gains in Performance? 47

1. Teacher's Background 48

2. Teacher's Approach and Program Emphasis 49

3. Teacher's Behavior in the Classroom 50

4. Classroom Activities 50

5. Classroom Equipment and Materials 51

6. Classroom Atmosphere 52

Which Program Inputs Are Associated with Specific Outcomes 52

1. Preschool Achievement 53

2. General Cognitive Ability 53

3. Learning Ability 53

4. Social Adjustment 54

Do Different Teacher and Program Approaches Influence Different Children in Different Ways? 54

What Kinds of Children Benefitted Most from Different Teacher and Program Approaches? 58

DISCUSSION 62

Factors Influencing Developmental Changes 62

Implications for Early Childhood Intervention 65

Summary 67

REFERENCES 68

APPENDICES

A - Measure Selection and Psychometric Characteristics 75

B - Selection of Variables for Analysis 84

BACKGROUND

INTRODUCTION TO THE STUDIES

In the Spring of 1966, the first prospective national evaluation of the effects of participation in full-year center-based Head Start programs was planned by the Research and Evaluation Office of Project Head Start. The evaluation was designed to identify what kind of program would be most effective for different children. In 1967 and 1968 the evaluation was continued and extended, with increasingly broad measurement of program characteristics and child development, and with designs that increasingly specified what kinds of program variations would be studied. The data were collected during these three years (1966, 1967, 1968) by a network of fourteen university-based evaluation and research centers, prepared for data analysis in 1969 by World Systems, Inc. (Contract OEO-B-99-5024) and independently analyzed during 1971-1972 by System Development Corporation (SDC, 1972a; SDC, 1972b) and Research Triangle Institute (RTI, 1972). The approach of having two teams independently address the same issue was adopted to reduce the controversy so often directed at the analytic methodology of major evaluation studies. The purpose of this report is to summarize the findings from these two technical reports on the immediate changes in child development associated with Head Start, and to present a statement of the extent to which these changes are enhanced or attenuated by different program experiences for different children.

The studies (i.e., the separate evaluations for 1966, 1967 and 1968) were designed to identify immediate changes associated with Head Start participation as well as the conditions under which these changes were greatest. There were no control groups of eligible children who did not attend Head Start since addressing the issue of what kinds of classroom experiences "work best" for different children, required comparisons within the Head Start sample and not between Head Start and non-Head Start preschool populations.

While the studies addressed only a limited array of the many questions asked about early education and Head Start, the data nevertheless represent a rich and varied source of information on the child in relation to his classroom experience. The questions asked were:

- What are Head Start children and families like and how do they differ from one another?
- How do background characteristics of Head Start families and children influence children's initial performance?

.What are Head Start programs like?

. What changes in children's social adjustment, motivation, scholastic readiness and cognitive performance are associated with participation in Head Start?

.What changes in parent attitudes, particularly toward education and society, are associated with children's participation in Head Start?

.How do background characteristics of Head Start families and children influence the gains children make in the program?

.How do teacher and program characteristics influence the gains children make in the program?

.What are the inter-relationships between different teacher and program approaches and benefits or gains for different kinds of children in Head Start?

OVERVIEW OF THE FINDINGS

1. Head Start children and families differed from one another on almost every dimension studied. This finding is consistent with reports by Bates (1972), Educational Testing Service (1971), and others of the heterogeneity of family circumstances and child development within "poor" families.

2. The family was the major determinant of initial performance differences among Head Start children. Specifically, parents of children who scored highest at the initial testing had higher aspirations and expectations for the child's education, were more accessible and read more frequently to him/her, seldom used physical punishment in disciplining the child and generally felt more confident about life.

3. Most Head Start programs were oriented in a nonspecific way to general personal-social development, but varied extensively from one another within that approach. This finding is consistent with the view of Head Start as a broadly-based and multi-aimed child development program as distinct from an intensive compensatory education program aimed at remediating cognitive deficits. It supports the position that Head Start is a multiplicity of programs rather than a single project or educational treatment.

4. Children gained from their Head Start experience; they gained more in specific skills and cognitive functioning (achievement motivation, adjustment to the testing situation, academic achievement) and less in basic cognitive abilities and other aspects of social development. The interpretation of these findings remains somewhat problematic. In certain areas of children's social development, e.g., peer sociability, it is not surprising that negligible gains were reported since Head Start children were already performing well in this area, leaving little room for improvement. However, in those areas where ample room for improvement did exist, i.e., basic cognitive abilities and cognitive functioning, the gains varied from slight to substantial. Two interpretations are plausible: first, that basic cognitive abilities are less easily modified by preschool interventions; or second, that the larger gains in motivation and achievement reflect the Head Start "whole child" program approach.

5. Parents' attitudes toward education became more favorable but parents' attitudes toward society did not change significantly. Parents' attitudes toward education were favorable initially and became even more favorable. However, no comparable improvements in parents' feeling of social alienation were brought about by children's participation in Head Start.

6. On the whole, child and family background characteristics were unrelated to magnitude of gains made during the program. Child's age and pretest IQ were the only characteristics consistently related to gains. Family characteristics, an important determinant of initial performance levels, did not appear to be an important determinant of changes associated with the program.

7. Teacher and program characteristics were the major determinants of magnitude of gains made in the program. All aspects of teacher and program inputs measured in these studies were related to gains in the areas of cognitive functioning, cognitive abilities and social adjustment. Generally speaking, cognitive gains were associated more with cognitive-oriented program emphases and classroom activities; social gains were associated more with social-oriented program emphases and with qualitative aspects of classroom atmosphere. However, the data do not indicate a clear division, within the "whole child" orientation of most Head Start programs, between specific combinations of program inputs and child development outcomes.

8. These data suggest that there was no one "best approach" for all children within the range of program variations studied. Two broad program clusters appeared effective for two different kinds of Head Start children: one cluster was related to the orderliness and structure of the environment and was particularly beneficial to children who initially lagged behind their peers; a second cluster was related to the social climate of the class and was more beneficial to children performing at higher levels upon entry into Head Start. The finding that what a child brings to a program influences how he is affected by it is consistent with earlier evidence from studies by Weikart (1969) and Miller (1971), and with the viewpoint articulated by the developmental theorist, J. McVicker Hunt (1961). These data are among the first large scale evaluations that report inter-relationships among program and child characteristics and, as such, offer tentative support for the policy of individualizing preschool experiences in Project Head Start.

HISTORY OF PROJECT HEAD START

Head Start is a comprehensive child development program designed to give children of poverty the services and experiences they need to help them develop as healthy, happy and competent individuals. Initiated in the summer of 1965 as part of the War on Poverty, Head Start was seen as helping other adult-oriented (e.g., Job Corps), service-oriented (e.g., Neighborhood Health Centers), or change-oriented (e.g., Legal Services), programs break the poverty cycle by early intervention. Since summer 1965, Head Start has served almost three million children in the eight-week summer programs and over one and one-half million children in the full-year programs.

Chief among the distinguishing characteristics of Head Start as conceived by the committee chaired by Dr. Robert Cook were:

- .provision of comprehensive services with particular attention to health and nutrition
- .emphasis on the importance of strengthening family life, and the ability of the parents to be the primary advocates, change agents, and educators for their children
- .focus on the child's motivational and social development and on the achievement of competence in everyday life including academic preparation for school

The report of the Cooke Committee (1965) reflects not only a belief in the value of early intervention but an awareness of the need for continuous comprehensive support if children of poverty are to reach their full potential. The Cooke Memorandum describes the seven major objectives of Project Head Start:

1. Improving the child's physical health and abilities.
2. Fostering the emotional and social development of the child by encouraging self-confidence, spontaneity, curiosity, and self-discipline.
3. Improving the child's mental processes and skills with particular attention to conceptual and verbal skills.
4. Establishing patterns of expectations of success for the child that will create a climate of confidence for his future learning efforts.
5. Increasing the child's capacity to relate positively to family members and others while strengthening the family's ability to relate positively to the child and his problems.

6. Developing in the child and his family a responsible attitude toward society, and fostering constructive opportunities for society to work together with the poor in solving their problems.

7. Increasing the sense of dignity and self-worth within the child and his family.

Head Start is often thought of in the context of compensatory education; that is, as a remedy for specific cognitive deficits in the participating children. Another common viewpoint sees Head Start as a program of early childhood education, designed to take advantage of the young child's rapid learning and development. A review of the Cooke Memorandum objectives, however, indicates that Head Start was planned as a broadly based program of early child development to deal comprehensively with the needs of young children in a context of community involvement and parent participation.

Within the framework of national guidelines specifying goals, program components, minimum staffing patterns, and policies governing eligibility and program administration (Head Start, 1967), individual Head Start programs have varied considerably in their goal emphases and have used a broad range of approaches including structured, skill-oriented training and cultural programs focused on the child's ethnic heritage. Not all Head Starts have been equally successful in the task of balancing the developmental expectations of the local cultural milieu and the developmental expectations of the larger community as expressed in requirements for school achievement. While approaches to the assessment of Head Start success have increasingly reflected the breadth of Head Start's objectives, assessment technology still lags behind new perspectives on cultural differences.

The origins of Head Start have been described in many reports (Hellmuth, 1967, 1970; Auleta, 1969; Anderson and Shane, 1971; Evans, 1971; Stearns, 1971; Braun and Edwards, 1972; Rowe, et.al., 1972). To summarize briefly, the inspiration for Head Start sprang from the following sources:

.Growing public and governmental support for efforts to reduce the obvious racial and economic inequalities in all parts of the country. This general concern led to the creation in 1964 of the Office of Economic Opportunity and to the declaration of the "War on Poverty."

- Increased recognition that educational inequalities are a major factor in creating and perpetuating inequalities in earning power and the general quality of living. 1/ One example of this recognition was the 1954 United States Supreme Court ruling on racial desegregation of the schools. In the ensuing efforts to redress racial inequalities, it soon became evident that many of the inequalities were a function of socio-economic differences. Attention then began to focus on how compensatory education for children from low-income families might be used to help break the poverty cycle.

- An accumulation of theory and evidence from early studies by the Iowa Child Welfare Group (Skeels and Dye, 1939) and from work by Hunt (1961) and Bloom (1964) showing that environmental factors in the early childhood years are particularly powerful in shaping children's future growth and development.

- Recognition of the disproportionate number of poor children with school problems and low achievement, and the hope that early education might reduce their incidence.

1/ The authors of the recently published study, Inequality: A Reassessment of the Effects of Family and Schooling in America (Jencks, et.al., 1972), argue that schools have little long-term effect on the socio-economic success of the individuals who attend them. From his reanalysis of the Coleman Report data, Project Talent's Longitudinal Study, and Census Bureau data on adult occupational status and income, Jencks concludes that differences or inequalities in educational opportunities and schooling are unrelated to socio-economic differences among graduates and draws from this correlational evidence the policy implication that school reform is important for improving the quality of the lives of children but contributes only marginally to reducing adult inequality and poverty.

.Research findings indicating that disadvantaged children are deprived of many of the opportunities which seem necessary for growth during the early years. 2/

In an attempt to take a first step towards ameliorating the negative effects of poverty on children, Project Head Start was launched as an eight-week pilot program in Spring, 1965. When Head Start attained instant popularity, enrollments exploded from 40,000 children in Spring, 1965, to over 500,000 children in Summer, 1965. Since it was recognized early (Hunt, 1966) that a longer program might be required to produce a more enduring and significant impact on children, families, and institutions, several Head Start programs were converted from eight-week pilots to full-year projects in August, 1965. However, few full-year programs actually became operational until 1966-1967 because of the problems involved in "starting-up" such an effort, e.g., locating suitable sites and staff.

In 1967-1968, almost 300 new full-year programs were funded and some experimental Head Start-related projects were initiated: Parent-Child Centers to provide services to toddlers (see Costello and Binstock, 1970), and Project Follow Through to continue providing services to Head Start graduates (see Bissel, 1971). Thus, in the years 1966-1968 the program had been initiated and rapidly expanded in all directions to serve increasing numbers of disadvantaged pre-school age children, and some younger and older disadvantaged children and their families as well.

2/ Since then, the existence of critical periods (i.e., times of maximum capacity for growth or susceptibility to irreversible damage), and the role of the preschool years as a critical period, have been challenged (Rohwer, 1970). Similarly, the relative value of child-focused (as contrasted to adult-focused) interventions has been questioned, based on the apparent failures of preschool, primary, and secondary age interventions (Averich and Donaldson, 1971; Mosteller and Stet, 1972). These recent reports do not deny the influence of the environment nor the importance of growth during the first six years of life; instead, they examine the relative costs and long-term benefits of different intervention approaches. In 1961-1965, when many of the first of these programs were initiated, however, intervention during the formative years was a logically clear mandate from the available data.

The 1968-1969 program year was the first during which Project Head Start could be considered to have achieved stable operations. This important historical fact provides some rationale for basing this summary report of 1966-1969 evaluation studies almost exclusively on the 1968 data.

It is also noteworthy that Head Start centers were not alone in their start-up year management problems. The Head Start regional Evaluation and Research Centers, scheduled to begin collecting national evaluation data in 1966, did not begin operations until September 1966 - January 1967. Since full-year Head Start classes began operation well before the fourteen Evaluation and Research Centers were ready, the data from 1966-1967 only approximates a "pre" and "post" study with an average interval between first and second testing of only four months. Fortunately, these start-up problems were minimized in the second and third evaluation years (1967 and 1968). The Evaluation and Research Centers continued operations until 1969, when another Head Start related project, Planned Variation (see Klein and Datta, 1972) was begun and the Centers were phased out.

These considerations, while rarely recognized in discussions of Head Start, are nevertheless significant in interpreting evaluation data from Head Start and similar programs as critical tests of the validity of conceptual positions on the effects of early intervention on child development.

It is now recognized (Bissell, 1971) that even well-planned, conceptually integrated, and highly focused programs require at least three years to implement on a modest scale; this suggests that while findings from the 1966-1969 evaluations of Head Start and similar studies can have substantial value for identifying likely immediate and long-range effects, they cannot individually, or as an average, 3/ be interpreted as conclusive or critical tests of the validity of Head Start as a conceptually based solution to a social problem.

3/ See Light and Smith (1971) for a discussion of issues in averaging results from several studies.

OVERVIEW OF HEAD START EVALUATIONS

Evaluation Issues. Initially, two issues received considerable attention in Head Start evaluations, including evaluations funded by Project Head Start as well as those undertaken by other agencies and individual researchers: (a) Is the program delivering the services it is obligated to deliver to eligible children and families? (b) Is the program of benefit, as expected, to children and their families?

The first issue focuses on the process of implementing a program and traces the stages of program development from the identification of a social need to the actual delivery of services and program impact. Addressing this issue enables researchers and program planners to identify where breakdowns in the process occur and to determine whether a program has been ineffective or merely improperly implemented.

The second issue focuses on outcome. Most outcome assessments do not distinguish between levels of program implementation; undesirable or negligible outcomes are often misinterpreted as indicators of program ineffectiveness rather than failure of program implementation. Unfortunately, in large scale evaluations which include many programs of varying quality, program implementation becomes a matter of degree, or a variable, that is usually not adequately controlled.

In addition to these broad issues concerning program effectiveness, three issues of a more substantive nature have gained in popularity since the beginning of Head Start and parallel the growing interest in the field of child development and early education. These issues are: (a) Does the success or failure of Head Start, when viewed solely as a compensatory education program, mean that compensatory education has succeeded or failed? (b) How durable are the benefits of early versus later school interventions? (c) What program, curriculum model, or approach works best?

Viewed in the context of these evaluation issues, the Head Start evaluation studies conducted in 1966-1969 and summarized in this report can be described primarily as outcome assessments of different program approaches within Head Start to determine which kinds of programs work best; the degree of program implementation was roughly ascertained on the basis of program compliance with Head Start guidelines. These studies were not designed to address the issues of the success of compensatory education or the durability of its effects; no control group data were collected and only immediate changes were measured.

Evaluation Studies. From the perspective of the type of research study conducted, Head Start evaluations can be grouped as follows:

1. descriptive studies of program compliance with Head Start guidelines;

2. one-site research and evaluation studies of individual Head Start programs;

3. national impact studies of the overall effects of the Head Start program where program characteristics are not linked to outcomes; and

4. national interactive model studies of specific program characteristics as they are related to outcomes.

1. Descriptive Studies. Between summer 1965 and summer 1970, at the request of Head Start, surveys of random samples of Head Start programs, centers, classes, staff, and children were conducted by the Bureau of the Census. The questionnaires were designed by Head Start evaluation staff and program specialists to assess compliance with national guidelines and to describe Head Start programs. A series of reports prepared by Bates (1969, 1970, 1972), Office of Child Development, provide an extensive description of each summer and full-year program since 1965. In general, the findings indicate compliance with Head Start guidelines, but also detail substantial variability along child, family, staff and program characteristics.

The Census surveys, in tracing the development of various program components over the history of Head Start, show that while most programs have been in compliance with guidelines, program quality has been (a) uneven between program administration and services components, and (b) uneven from project and even from class to class. Variations seem, predictably, to be greater in the early programs than in more recent years; trend analyses have shown steady progress in many areas of significance to Head Start, such as provision of extensive in-service training and election of parents to Policy Councils.

2. One-site Research and Evaluation Studies. Many of these single-project studies were funded by Head Start to assess the immediate effects of program participation on child development and, in some cases, to follow the children after they entered public school.[4/]

4/ Most of these reports are available through the ERIC Early Childhood Education Clearinghouse, University of Illinois, Urbana, Illinois

It should be noted that the research designs for Head Start evaluations typically are similar in methodological features to the research studies of experimental preschool programs, few of which permit drawing unambiguous inferences regarding program effects. Nonetheless, some of these single project evaluations do involve control groups of non-participating children, and many use pre and post measurements. 5/

In general, most intervention programs have been found to produce a statistically significant, immediate effect on cognitive and linguistic performance. These gains have been greatest when (a) the program was directed toward specific educational objectives, as in many "experimental" Head Start programs, and (b) the children participated for a nine-month school year rather than a three-month summer session.

Compared to the substantial documentation of cognitive and linguistic gains, data pertaining to the effects of early intervention on social-emotional development are sparse. This is all the more regrettable since "increasing the self-concept", for example, is a nearly universal objective of Head Start programs. One major problem is that although there are no entirely satisfactory measures of cognitive, perceptual-motor, linguistic, and academic skills for preschoolers there are even fewer psychometrically acceptable measures of personal and social growth (see Hoefner, et.al., 1971; also see Walker, 1974). However, what evidence exists is positive (see Grotberg, 1969); there are numerous observational and anecdotal accounts of the effects of preschool interventions on improving self-confidence, peer relationships, interactions with adults, curiosity, and other personal-social characteristics of disadvantaged preschool age children.

With due regard for methodological and other limit to present knowledge, most researchers conclude that immediate changes in the development of low-inome preschool children are attributable to preschool intervention programs, outcomes which are within the current state-of-the-art for small scale programs and are theoretically possible in larger scale programs.

5/ The findings from the Head Start one-site research studies have been reviewed at length by Grotberg (1969), Datta (1969), Stearns (1971), Butler (1970), and others.

3. National Impact Studies. Only one national impact study has been reported by Head Start where the objective has been to assess performance changes without assessing the specific conditions under which such changes occur. Known as the Westinghouse Report (Cicarelli, et.al., 1969), this study was funded by the OEO Division of Research, Program Planning, and Evaluation in Summer, 1968. The design compared the performance of Head Start and non-Head Start children in 104 sites who were attending first, second, and third grades in October-November, 1968. About two-thirds of the Head Start children had previously attended summer programs and one-third had previously attended full-year programs. Although aspects of the study are considered methodologically controversial (Smith and Bissell, 1970; Campbell and Erlenbacher, 1970), the findings reported in the Westinghouse Report are similar to those of one-site studies discussed earlier; children tested in first grade soon after leaving full-year Head Start achieved higher scores on cognitive measures than did non-Head Start controls; however, by second or third grade Head Start children on these cognitive measures. Moreover, the Westinghouse Report found no benefits, either cognitive or motivational, associated with summer attendance in Head Start.

While the findings of the Westinghouse Report were similar to those of previously described one-site studies, the conclusions drawn were not. Most researchers interpreting similar findings concluded that Head Start programs should be extended through primary school, reasoning that continued intervention would help to sustain the effects. Cicarelli, et.al., concluded however, that most Head Start programs did not produce the expected outcomes and recommended curtailing the projects and modifying the full-year Head Start program by installing known successful preschool education models. An experimental approach to Head Start was strongly urged, with further program expansion to serve more children suspended until proof of the effectiveness of these changes became available (note that Head Start serves only approximately 11% of children eligible as defined by the age and poverty criteria).

4. National Interaction Model Studies. Most national evaluations funded by Head Start have been based on an interactive assumption: the belief that different programs have different effects on different children. Prompted in part by Hunt's description of the importance of the proper "match" between the child's competencies and the program's challenges, and in part by the diversity of Head Start children which seemed too great to be quality well served by one approach, this belief has been shared by such researchers as DiLorenzo and Salter (1969), Karnes (1969), and Miller (1971) who have investigated the different effects of several well-known curricula.

Head Start national evaluations have included: a review of the summer 1965 reports (Planning Research Corporation, 1966); an analysis of the associations of the first full year (1966) program experience

(Planning Research Corporation, 1967); an evaluation of the Summer, 1966, Head Start programs (Boyd, 1966); a study of the effects of Head Start on community institutions (Kirschner Associates, Klein and Datta, 1971); and the Educational Testing Service Longitudinal Study of Disadvantaged Children and their First School Experiences (Shipman, 1970). From 1966 to 1969, the national evaluation program was undertaken by the 14 Head Start Evaluation and Research Centers, (see Table 1). It should be noted that in all of these studies assessments of program and teacher characteristics and of classroom process variables have received as much attention as assessment of initial and final levels of performance. Analyses of these evaluations, including the 1966-1969 studies, have been directed at identifying the conditions associated with greatest gain for different children.

Yet, despite the widespread belief in the interactive assumption and commitment to investigating its validity for preschool intervention, there has been to date relatively little direct evidence from Head Start data showing that child, family or even program characteristics are differentially associated with the magnitude of gains made in the program. While in every study some differential effects and interactions have been reported, their size and relative infrequency is more striking than their presence. But these earlier data in particular are fraught with methodological problems which likely obscurred interactions, e.g., lack of interactions can be attributed to the relatively restricted range of classroom characteristics in the 1965 and 1966 programs, and/or to the restricted range of outcome variables (nearly all of them academically oriented) which were studied.

As a comprehensive child development program, of which classroom experience is but one key component, Head Start has been influenced by both educational and developmental aspects of research on the effects of early intervention. The nature of the Head Start program -- broadly child development oriented in its objectives, and, in the case of the classroom component, more narrowly education-oriented -- suggests that at all decision making levels the interplay or interaction between classroom/teacher and child characteristics must be taken into account.

The need to assess interactions as sensitively as possible has been reflected implicitly in the evolution of Head Start national evaluations from the simple random sample designs of Summer 1966 to the planned variations and interventions of 1969, 1970, and 1971. The evaluation studies summarized in the present report represent an intermediate stage in this evolution. They do not provide the most direct and sensitive test of program-child interactions; nor do they, with the exception of the 1968-1969 data, sample a wide range of program variations. Nonetheless, the 1966-1969 evaluation studies were directed, to as great an extent as possible, to what was identified as one of the central issues in early childhood intervention: What programs work best and for what kinds of children.

TABLE 1

HEAD START: NATIONAL RESEARCH AND EVALUATION STUDIES, 1965 - 1972

YEAR	PROJECT
Summer 1965	Local research studies; PRC report; Census Study
Summer 1966 Full Year 1966-1967	ETS national evaluation, Census survey E&R Center study; Census survey; PRC report
Summer 1967 Full Year 1967-1968	Local research studies; Census survey E&R Center study; Census survey
Summer 1968 Full Year 1968-1969	Local research; Census survey E&R Center study; Census survey; ETS longitudinal study, Year 1; (Westinghouse report, Spring 1969)
Summer 1969 Full Year 1969-1970	Census survey Planned Variation, Year 1; ETS longitudinal study, Year 2; Census survey; community impact study
Summer 1970 Full Year 1970-1971	Census survey Planned Variation, Year 2; ETS longitudinal study, Year 3; Census survey
Summer 1971 Full Year 1971-1972	Health Start, Year 1 Planned Variation, Year 3; Home Start, Year 1; ETS longitudinal study, Year 4; parent participation study; staff mobility study
Summer 1972 Full Year 1972-1973	Health Start, Year 2 Home Start, Year 2; ETS longitudinal study, Year 5; others in process

METHODOLOGICAL ISSUES

Evaluation methodology has changed rapidly since 1965 when the first Head Start programs were founded. So rapid are these changes that in the two or three year time span between initiating a study and preparing a final report, the research design may have become obsolete.

Examined from this evolving methodological perspective, Head Start program evaluations designed in 1970-1973 are likely to differ from those designed in earlier years in three major respects:

- more realistic estimates of the length of time required to implement a new program or program objective. Recent evaluations of new Head Start programs such as Health Start (Urban Institute; final report completed 1973) and Home Start (High Scope Foundation; final report due 1975) focus the first year evaluation activity on assessing management and administration including staff selection and training and then, during the second or third year of the program, assess impact on the target population.

- more closely linked to program objectives. Today's evaluators work directly with program planners and managers to identify appropriate criteria of program success and realistic dates when short-term and long-term effects can be expected. Today's evaluators are more conscious of the need for developing a repertoire of sensitive and reliable outcome measures or at least alerting program planners when no satisfactory outcome measures are available to assess specific program objectives. Recent evaluations are thus more likely to gather data by direct measurement of specific behaviors that are expected to be influenced by the program and less likely to draw upon data from indirect measures of general traits or abilities.

- more rigorously experimental in design. The limitations of statistical manipulations of data to compensate for playing fast and loose with the design have become more apparent in recent years. The problem is not primarily one of assessing change; assessing change is a relatively straightforward matter of comparing initial with later measurements. The problem is one of ascribing change unambiguously to the program intervention or ascribing relative change to different programs. In order to conclude that an outcome actually resulted from the program input, evaluation theorists argue that certain minimal design criteria must be upheld, even at the expense of scaling-down national studies.

The studies summarized in this report were designed with some, but not perfect, foresight into these developments. Among the limitations are:

1. Lack of a control group of children who have not attended Head Start.

2. Non-random assignment of children to different program types within Head Start.

3. Non-uniformity of time and conditions of data collection.

On the other hand, the advantages of these data include:

1. Extensive descriptions of what actually happened in the classrooms.

2. Data on child, family, and program characteristics for large national samples for three consecutive years.

3. Assessment of personal and social as well as academic and cognitive development.

4. Measurement at two points in time -- roughly the beginning and end of the program.

DESIGN ISSUES

There are two design issues addressed in the 1966-1969 evaluation studies which are especially significant for early childhood intervention research: the measurement of social-emotional development and the measurement of program-child interactions.

Measuring social-emotional development. Large-scale studies of child development programs require tests which are easily administered and scored, with high statistical reliability and validity. Unfortunately, there are few measures of young children's social-emotional development which meet these criteria. The best data in these non-cognitive areas come not from standardized tests, but from controlled time-sampling of children's behavior and clinical observations of children. For example, it has been argued convincingly by Zigler and Butterfield (1968) that changes in personal-social attitudes rather than changes in actual cognitive abilities account for the small increases in IQ test scores produced by early intervention programs. Zigler and Butterfield maintain that social and motivational factors can be readily influenced by early childhood interventions and that these changes in turn improve performance in the testing situation. More recent studies by Emmerich (1971) and Dittman, et.al. (1971) using

observational techniques, lend support to this argument by reporting substantial socialization gains for Head Start children during the first half of the program.

A common misconception is that Head Start evaluations and other research studies have neglected to collect personal-social data or to develop new measures. If anything, however, more has been invested in this area and in studies of ongoing classroom activities than in the measurement of cognitive and linguistic changes. With respect to national evaluations conducted in 1966-1969, a variety of rating scales of social-personal development were prepared. These included not only a variety of children's behavior rating-scales to be used by teachers and testers, but experimental tests of achievement motivation and extensive observations of children's unstructured social interactions with older children and adults. The data have proved expensive to plan for, to collect and regrettably, to analyze. This summary report, in addition to the final reports prepared by System Development Corporation and Research Triangle Institute 6/ represent analyses of only a fraction of the total data banks resulting from the 1966-1969 evaluations. Subsequent analyses of these data are likely to prove a fruitful source of new knowledge in the area of personal-social development.

Measuring interactions between child and program characteristics. The 1966-1969 evaluations were designed to test the "interactive hypothesis" or interactive approach to early child development programs: that some kinds of program experience may be better than others for certain children and that what is the best program for one child may not be the best for another. The most direct test of the interactive hypothesis would require a comparison between individualized programs (where children's needs/competencies are matched to curriculum) and undifferentiated programs (where all children receive the same curriculum). If the interactive hypothesis were substantiated, more children would be expected to progress in individualized programs than in the undifferentiated programs.

The 1966-1969 evaluations provide an indirect test of the interactive hypothesis; children were not matched to curriculum. However, the design did provide a systematic search for (a) overall changes, (b) program characteristics differentially associated with changes, (c) child characteristics differentially associated with changes, and (d) patterns of child/family and teacher/program characteristics differentially associated with changes.

6/ Final reports prepared by System Development Corporation and Research Triangle Institute on the evaluation of Head Start from 1966-1969 are filed in ERIC/ECE (Early Childhood Education Clearinghouse, in Urbana, Illinois).

DATA COLLECTION

The evaluation data were collected in 1966-1967, 1967-1968, and 1968-1969 by 14 regional university-based Evaluation and Research Centers. 7/ The evaluation designs and types of measures developed during the three years of evaluations are described below and are summarized in Table 2 (see also Datta, 1969).

Evaluation data for 1966-1967. Head Start classes were selected for study by the Centers to provide a representative sample of Head Start programs. Programs were selected for variability on the following characteristics: type of delegated agency operating the Head Start program (Community Action Agency or Local Educational Agency); geographic location (rural or urban); ethnic backgound of the child (Black, White, Spanish-surnamed, Polynesian). The Evaluation and Research Centers studied as many Head Start classes as possible within the confines of their geographic areas, drawing samples of six to eight children from each class.

Evaluation data for 1967-1968. Head Start classes were selected for study by the Centers to provide a more differentiated sample of Head Start classroom activity, but not necessarily a sample that was repsentative of Head Start programs. The purpose of this sampling modification was to provide a greater opportunity to assess differential effects within a classroom, i.e., interactions between program and child characteristics. This more differentiated sample of classroom activity was achieved in two ways: first, classes were selected predominantly on the basis of classroom style and process characteristics, although data were also gathered on those characteristics identified in the 1966 evaluation design; and second, all children within each class were included in the sample. 8/

7/ Head Start Evaluation and Research Centers were in operation from 1966-1969. They were established by the Head Start National Office for the purpose of developing and selecting, with guidance from the Head Start National Office staff, the designs and measures for the 1966-1969 evaluation studies. Data were then collected and analyzed by the Centers under the direction of the national office. It should be noted that the beginning of full-year Head Start classes in 1966 did not overlap completely with the operation of the Evaluation and Research Centers. As a result, the 1966-1967 data do not provide adequate "pre" and "post" assessment--the average interval between initial and final testing was only four months and some children were enrolled as long as seven months before they were first tested.

8/ In the 1967-1968 evaluation study, extensive data were collected on classroom processes, social-motivation and perceptual-cognitive development. Much of these data remain to be analyzed and are referred to collectively as "cluster data", indicating how these particular data were collected by four clusters of Evaluation and Research Centers.

TABLE 2

HEAD START NATIONAL EVALUATION DESIGNS AND MEASURES 1966, 1967 and 1968

	1966	1967	1968
Design			
N Classes studied	225	177	148
N children tested	1806	1889	1989
Basis for selecting classes	variability by categories of program child characteristics	variability by classroom style and process characteristics	variability by experimental interventions
Selection of children	random within class	all children in each class	no prior H/S experience
Common Variables Across Years			
Classroom activity	observer form (ORF)	observations of classroom activity (OSCI I)	observations of classroom activity (OSCI II); rating of teacher behavior (POT); teacher questionnaire
Teacher characteristics	questionnaire	questionnaire	questionnaire
Class resources and facilities	observer form	observer form	observer form
Family background	interview	interview	interview
Administration	--	--	checklist
Cognitive performance	IQ test (Binet)	IQ test (Binet)	IQ test (Binet)

TABLE 2 (cont.)

	1966	1967	1968
Achievement	test instrument (PSI)	--	test instrument (PSI)
Motivation	teacher rating	tester rating	tester rating; experimental measure (GUMP)
Social Adjustment	tester rating	observation of children's social interactions (SIO)	sociometric instrument
Cognitive Style	--	--	rating instrument (Hertzig-Birch)
Data Collection			
Mean week of Pretest	12	--	7
Mean week of Post-test	28	--	32
Mean interval of weeks	16	--	25
% with prior Head Start	30%	15%	29%

Evaluation data for 1968-1969. Many Evaluation and Research Center directors maintained that sampling on the basis of the "natural variation" in classroom activity, as in the 1967 study, included too few sites that were examplars of new approaches in early childhood education. Moreover, many directors pointed out that the degree of program effectiveness, in terms of child development outcomes, was not uniformly high across all Head Start programs. Thus in 1968, each Evaluation and Research Center director developed a research design which compared "regular" Head Start classes (reflecting an experimental and general approach) with "intervention classes" (reflecting an experimental approach or special educational emphasis). The "interventions" undertaken were diverse and included projects emphasizing language, basic cognitive processes, parent education, community participation in the setting of program objectives and physical facilities. The particular types of intervention included in the 1968 evaluation study are summarized in Table 3. It was reasoned that an evaluation design comparing "regular" with "intervention" Head Start classes would permit drawing clearer inferences about differential program effects and program-child interactions, at the expense of a less programmatically-representative sample of Head Start classes. In the 1968 study there were no additional sampling requirements other than selecting children with no more than four months of prior Head Start experience and maintaining the geographic distribution which was the natural result of the location of the Centers. Most Center directors chose sites relatively near their universities which would maintain the desired balance of rural and urban settings.

MEASURES

In every program evaluation to date, an attempt has been made to collect data relevant to the several Head Start objectives; the attempt has been more successful with parent and class characteristics than with child characteristics. It has proved difficult to find or construct reliable, easily administered, readily interpretable measures which are also psychologically meaningful for different age and cultural groups. In many studies, the evaluator must choose between investing in (a) untried measures that look meaningful but may prove to be costly, insensitive, or unreliable, and (b) standardized measures that are inappropriate developmentally or culturally, and may prove to be uninterpretable or misleading because the usual interpretation is invalid for the Head Start population. The compromise of the 1966-1969 evaluation studies was to use both new and standard instruments, and to interpret with care.

For purposes of this summary report, the data from the 1966-1967 and 1967-1968 evaluation studies will be used selectively only to indicate replication of important findings obtained in the 1968-1969 data or to highlight a theoretically meaningful point. Detailed descriptions of the variables and measures used in these evaluation studies are

TABLE 3

EVALUATION AND RESEARCH CENTER EVALUATION DESIGNS FOR 1968 - 1969

Center[1]	N Classes[3] R	I	Objective of Intervention or Study
Boston	4	8	Effects of community participation
Syracuse	14	-	Teacher praise/blame behavior
Bank Street	8	8	Comparison of Bank Street and "structured" HS
Temple	6	6	Effects of greatly enriched materials
South Carolina	2	12	Language development program[2]
Tulane	2	10	Language development program[2]
Southern	1	11	Language/parent education[2]
Texas	2	11	Language development program[2]
Kansas	4	7	Behavior modification training
Michigan	-	9	Piagetian training-cognition
UCLA	8	10	Teacher/parent goal setting
Hawaii	-	8	Language/parent education[2]

1/ The University of Chicago (Dr. V. Shipman) and Teachers' College (Dr. R. Thorndike) participated in the 1966-1967 and 1967-1968 studies only.

2/ Tulane, South Carolina and Texas had the same design and program; Hawaii/Southern also had the same design and program.

3/ Intervention (I) Head Start classes receiving experimental intervention programs were compared with Regular (R) Head Start classes receiving no experimental intervention.

summarized in Table 2 and Appendix A. Since this summary report is based predominantly on the 1968-1969 data, the characteristics measured in 1968-1969 are described in the following section and explained more thoroughly in Appendix A.

Characteristics and measures used in the 1968-1969 evaluation study. The 1968-1969 data include information on children and families as well as teachers and programs. The data were collected using a total of 24 forms, questionnaires, or test instruments. Additional demographic information (e.g., age, sex, urbanicity, ethnic background) was already available for all children on a master data card.

1. Cognitive Development Characteristics

   Intelligence. The Stanford-Binet IQ test, a commercially available, standardized, and widely used instrument, was employed as a measure of general intelligence. Scores were interpreted as reflecting a combination of experience, learning ability, and motivation. As a measure of performance, and not necessarily an accurate measure of underlying capacity, Stanford-Binet scores have been found to be sensitive to a variety of educational interventions and predictive of later school achievement.

   Ability to learn a new task. In 1968-1969 this variable was assessed by the subscale of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI). The subscale, entitled Animal House, is a relatively culture-free and non-verbal measure of this aspect of cognitive functioning.

   Achievement and school readiness. One aspect of cognitive functioning which is not tapped by standardized IQ tests such as the Stanford-Binet is cognitive functioning at the more concrete level. The Caldwell-Soule Preschool Inventory was used as a measure of achievement and school readiness. It was designed to be especially sensitive to the cognitive changes expected from preschool intervention programs.

2. Social-Emotional and Motivational Characteristics

   Adjustment to a novel (testing) situation. When a standardized test such as the Stanford-Binet is administered to a child, the child's behavior during the testing session, independent of his performance on the test, can be considered an important index of certain aspects of child development. The Inventory of Factors Affecting the Stanford-Binet is a rating instrument that was used by examiners to assess children's behavior during the test. The ratings are interpreted as reflecting traits such as impulsivity, concentration, and activity levels.

Cognitive style. Cognitive style is yet another aspect of child development that can be inferred from the child's behavior during the Stanford-Binet testing session. The Hertzig-Birch Response Style instrument was designed and used to measure various cognitive styles that the child employs when responding to test items. For example, a distinction is made between an incorrect response where the child has "worked at" a problem solution and an incorrect response where he has not.

Achievement motivation. Achievement motivation, an important component of actual shool achievement, was assessed by means of an experimental measure (Gumpgookies) that required the child to choose between two pillow-like creatures, one of whom is engaged in achievement oriented behaviors and one of whom is not.

Children's sociability. Children's relative popularity or isolation in the classroom, as rated by other children, was assessed by means of a Sociometric Play Situation instrument. Because of difficulties encountered in scoring the test, only social isolation scores were included in the data analysis.

3. Parent and Family Characteristics

Socio-economic status, parental attitudes, styles of parent-child interactions, and household routines. A structured Parent Interview with parents (usually the mother) was given at the beginning and at the end of the program for the purpose of collecting information on these characteristics.

4. Teacher and Program Characteristics

Data on teacher and program characteristics came from two types of measures: self-reports by the teachers and objective classroom observations by trained observers.

Characteristics measured by means of self-reports included:

Program emphasis, goals, and teaching approach. Teachers provided information on these characteristics in a structured Teacher Interview.

Administrative variables. Teachers provided information on a host of administrative related variables by completing a Checklist of Administrative Variables.

Teacher's background. Each teacher completed a questionnaire (Characteristics of Teaching Staff Form) which requested

demographic, educational and experiential background information about the teacher and teacher aides in each classroom.

Program facilities and resources. Teachers provided information about the materials available in the Head Start centers, along with details of the daily schedule and yearly calendar, on an instrument called the Class Facilities and Resources Inventory.

Characteristics measured by means of observational techniques included:

Teacher's behavior in the classroom. A trained observer rated each teacher on 33 five-point scales (Post Observation Teacher Rating Scale) reflecting behaviors such as reliance on on-going classroom activities.

Teacher's and children's behaviors in the classroom. An observational instrument (Observation of Substantive Curricular Input) consisting of a complex time schedule was used by trained observers to record the behavior of both children and the teachers in the Head Start classroom. Among the most important codes were those reflecting the context and context of the children's activity, the form of teacher control used, and the extent of teacher involvement in the activity.9/

STATISTICAL ANALYTIC MODELS

Two contractors (Research Triangle Institute, RTI, and System Development Corporation, SDC) were selected for the purpose of analyzing these evaluation data; both were provided with identical data files. The contractors were instructed, however, to make independent decisions about the selection of variables and choice of analytic techniques. As would be expected, the contractors differed in their decisions. A complete description of these differences is presented in Appendix B, but two important differences in their approaches are worthy of mention here.

First, although both contractors used a two-stage process for identifying and selecting a workable set of variables to include in the statistical analyses, the sequence of the stages differed. Research Triangle Institute selected variables first on a rational basis and then

9/ Factor analyses of these data identified six teacher and classroom variables to be utilized in the subsequent statistical analyses. These variables are identified and briefly described in Appendix A.

on a statistical basis: 90 variables that appeared relevant to outcomes were selected and then screened for psychometric shortcomings (missing data, low variability, low internal consistency and high redundancy). System Development Corporation, on the other hand, selected variables first using a statistical criterion and then a rational criterion: 203 variables were selected on the basis of changes in correlation with pre and post scores on performance variables; then other variables were added which appeared to relate, hypothetically, to the questions addressed in the study.

It should be noted that RTI and SDC used similar methods to assess the influence of child, family, and program characteristics on performance measures -- i.e., "gain scores", the difference between pre and post measurement, adjusted for pre score level. However, the two contractors adopted very different methods for examining the interactions between child/family and program characteristics.

RTI analyzed gain scores in the context of several child and program factors simultaneously. For example, RTI analyzed Stanford-Binet gain scores as function of child's geographic region, his pre score on the Stanford-Binet, his age, and the extent of program input or classroom structure. Other analyses took into account factors such as the child's gender, socio-economic status of the family, length of pre test - post test interval, and different programmatic characteristics. This method is maximally sensitive to the effects of a particular variable when the effects of all the other influencing variables are taken into account. As a statistical approach, it is especially useful for singling out the relative effects of several interdependent ("confounded") influencing variables. At the same time, however, it is quite insensitive to the interactions between such variables, especially interactions of small magnitude.

In contrast, the SDC analyses were less cautious about interdependency among influencing variables but more sensitive to small interactions. Rather than take several factors into account simultaneously, SDC computed for each program variable, separate analyses for subgroups of children broken-out by age (younger or older), pre-test IQ (high, middle, or low), urbanicity (urban or non-urban), and region (South or non-South). Interactions were inferred if relationships between program variables and outcomes occurred for one subgroup and not for the other, or if the direction of the relationship between program variables and outcomes was different for different subgroups.10/

10/ SDC's procedure resulted in over 100 one-way analyses of variance computed separately for each influencing variable. It is probable on the basis of chance that the results of some of these analyses are spurious. RTI's procedure, by comparison, consisted of a series of four-way analyses of variance which examined gain scores as a function of several influencing variables simultaneously. It is less probable that the statistical interactions reported by RTI are spurious; but it is possible that some true interactions may have gone undetected in this model.

Neither of the two approaches is more correct than the other. If the results are different, it is because slightly different analytical questions are being asked by the two models. The juxtaposition of the two approaches in this report should serve as a reminder that, computers nothwithstanding, data analysis remains much a matter of judgment.

FINDINGS

The findings reported here are based on the 1968-1969 evaluation data; results from the 1967-1968 and 1966-1967 data are indicated only where parallel variables exist and meaningful comparisons with the 1968-1969 data can be made. In general, the statistical analyses undertaken by System Development Corporation and Research Triangle Institute report similar findings and lead to similar conclusions. However, important differences between the SDC and RTI analyses do exist and are noted and explained when possible. For purposes of this summary report, only very reliable, statistically significant ($p < .01$) findings are presented except when the psychological significance of a particular variable or relationship warrants attention to less reliable (i.e., $p < .05$) findings. The findings are organized according to the particular substantive question they address, as determined by the overall design of these evaluation studies.

ENTERING CHARACTERISTICS AND PRESCORE PERFORMANCE

What Were the Head Start Children and Families Like?

The children included in the 1968-1969 study ranged from two and one-half to six years of age at the time of initial testing. About 60% were preschool age (48 to 59 months) and about 22% were kindergarten age (over 60 months). About one-third lived in the South (34%) and over three-quarters lived in cities with populations over 50,000. Many ethnic groups were represented, but the largest proportions were Black (68%), White (18%), Mexican-American (6%), and Polynesian (4%). The sample was evenly divided between boys and girls (51% and 49% respectively). Compared to the sample of children in 1967-1968, the children in the 1968-1969 study were a little older, more urban, more Southern, and more Black. Compared to the population of all Head Start children, as described in Head Start census surveys, the sample underrepresented older children and over-represented Black children.

On the whole, Head Start children came from families with low incomes and little education. The median family income in 1968-1969 was slightly above $4000 per year for the 6.7 persons in each family (the corresponding figure for all Head Start families is $3210). Still the per capita income of the Head Start children was considerably below the national average (about $597 as opposed to $2,213). About 60% of the parents (mothers and fathers) had not completed high school. About 60% of the mothers were unemployed; of the remainder, most were unskilled or semi-skilled workers (13% and 19% respectively). Only 11% of the fathers (living at home) were unemployed, the majority holding unskilled or semi-skilled jobs (38% and 28%). Most revealing about

the socio-economic status of the Head Start families is the fact that 61% of the families relied on parental earnings and were still eligible for Head Start -- even two "breadwinners" did not keep families above the poverty live. Sub-employment, rather than unemployment, accounted for the poverty of most Head Start families. Over one-half of the children (55%) lived with both parents, but nearly one-third (29%) lived with only their mothers. The levels of parental education and employment, and the proportion of children residing in two-parent families were essentially the same in 1967-1968.

The parents of Head Start children began the program with overwhelmingly favorable attitudes toward it -- 89% had the highest possible score on this scale. Their views about the value of education were also positive, but in the middle of the range of possible scores. On the other hand, Head Start parents did not reveal strong feelings of personal power over their environment or their children's schools nor a high degree of involvement in the community; overall they demonstrated moderately strong feelings of alienation. In 1967-1968 parents' attitudes were equally positive toward Head Start (92% indicated they would send another child to the program) but slightly less positive toward education in general.

When interviewed shortly after the 1968-1969 Head Start year began, mothers reported a marked difference between their aspirations and realistic expectations for their child's future. About half of the parents had high aspirations for their child's career and education (college degree), but only 14% had correspondingly high expectations for actualizing their goals. In 1967-1968, by comparison, mothers' aspirations were slightly lower (80% aspired to some post-high school education for their child) and the discrepancy between their aspirations and expectations was slightly greater (30% expected to realize their goals).

Summary. The 1968-1969 sample of Head Start children varied in age, sex, geographic location and ethnicity but was not completely representative of the total Head Start population: fewer older children and more Black children were included in the sample. Head Start children came from poor families that were below the national average in income and education and who suffered more from sub-employment than unemployment. Although the parents had favorable attitudes toward the program and education in general, they felt little personal power over their environment including the schools. Consistent with their feelings of alienation, Head Start mothers revealed a substantial discrepancy between aspirations and expectations for their children's education.

Were Differences Among Children's Prescores Related to Background Characteristics of the Children and Families?

At the time of initial testing, children's scores varied on

cognitive and emotional outcome measures (described in the preceding chapter). The question then arose whether the differences were systematically related to any background characteristics of the children or their families. Several statistically reliable and substantial relationships between the child/family characteristids and prescores were indeed found. These relationships express only association between the entering characteristics and the prescores; they do not indicate causality.

With the exception of the influence of prior Head Start experience, all of the relationships presented in this section are taken from the RTI analyses. Unless otherwise noted, comparable data from 1967-1968 were not available in the two reports.

At the time of initial testing, differences in children's prescores were systematically related to certain entering characteristics. These relationships are as follows:

1. Younger children performed better on cognitive tasks but appeared less well-adjusted socially. Younger children compared with older children had higher scores on measures of intelligence, ability to learn a new task, and achievement motivation. However, they also showed more behavior and motivation problems and greater feelings of inadequacy.[11]

2. Children living with only their mothers showed more behavior problems. Children from single-parent families, compared with children from two-parent families, showed a greater incidence of behavior problems in the testing situation, but they did not differ along other dimensions of social adjustment -- i.e., feelings of inadequacy and motivational problems.

3. Children with high aspiring mothers performed better on cognitive tasks. The mother's level of aspiration for her child's education was consistently and positively related to the child's prescores on measures of intelligence, school readiness, ability to learn a new task, and achievement motivation. On the Stanford-Binet IQ test, for example, children with high aspiring mothers scored seven points higher than children with low aspiring mothers (89 vs 82 respectively). All of these relationships were positive and were found in both the 1968-1969 and 1967-1968 data. It should also be noted that similar patterns were found with respect to the mother's educational expectations for her child, again in both sets of data.

11/ Younger children were not distributed evenly across all geographic regions. Thus the pattern described for younger children may well be the result of other factors associated with (and not separable from) younger children in this study, specifically, geographic region.

4. Children with optimistic parents performed better on cognitive tasks. Children of parents who were relatively less pessimistic about life scored higher on measures of intelligence, school readiness, and the ability to learn a new task. The difference on the Stanford-Binet was especially noteworthy: children of parents rated low in pessimism scored 8.4 points above children whose parents rated high in pessimism.

5. Children who were frequently read to performed better on cognitive tasks. Children who were read to frequently by adults and who had relatively easy access to adults tended to have higher prescores on measures of school readiness. The difference was not statistically significant for the factor of accessibility but was both statistically significant and substantial for frequency of reading: children who were "seldom or never read to" obtained scores approximately eleven points below scores of children whose parents read to them "very frequently".

6. Children whose parents seldom used physical punishment performed better on cognitive tasks. Children whose parents made little use of physical punishment obtained Stanford-Binet prescores 9.5 points above those obtained by children whose parents reported using severe physical punishment. Curiously, this relationship occurred only on measures of cognitive performance, and not social-emotional development.

7. Children with prior Head Start experience performed better on cognitive and social-emotional tasks. About 20% of the children studied had had prior experience in Head Start (including full year preschool or Head Start, summer Head Start, or less-than-full-year Head Start programs). Previous Head Start experience was positively related to prescores on several cognitive variables (higher scores on measures of intelligence, school readiness, cognitive style, achievement motivation) and social-emotional variables (fewer motivation problems and less frequent feelings of inadequacy.) However, there was no difference between children who had had previous Head Start experience and those who had not on either the incidence of behavior problems or social isolation.

The effects of prior Head Start were only partially replicated in the 1967-1968 data: children with prior Head Start were more verbal and engaged in more social interactions than children without such experience, but they were also less well-adjusted in the testing situation. No difference occurred on measures of general intelligence. The disparity between the 1968-1969 and 1967-1968 data is probably accounted for by the longer average exposure to Head Start among "experienced" children in 1968-1969.

It should also be mentioned that some inconsistent relationships between children's ethnicity and prescores emerged but will not be surmarized here since no meaningful interpretations can be made. These patterns were not stable across measures, did not replicate across years,

i.e., the 1967-1968 data, and were confounded with several other factors (region, age and socio-economic status).

Summary. Overall, family background characteristics were more closely related to children's prescore performance than children's background characteristics. Whether a child was younger or older, male or female, Black, White or other, had less bearing on the child's entering performance than did the family background of that child. Specifically, as indicated in these data, children with the highest prescores overall (those demonstrating most competence upon entry) came from families where parental aspirations and expectations were high, the mother felt confident about life, adults were accessible, the child was read to frequently, and little physical punishment was used.

It is noteworthy that children who had had some previous experience in Head Start obtained higher scores upon entry in 1968, suggesting that their initial exposures to Head Start did have some positive impact.

What Were the Head Start Programs Like?

The Head Start classes were described in terms of non-substantive program characteristics and substantive classroom-orientation characteristics. The 1967-1968 data provide the most complete description of program characteristics; the 1968-1969 data best describe classroom-orientation characteristics.

Program characteristics. Most Head Start programs (91%) operated five days a week, many (65%) for three to four hours a day. The majority of classes (78%) had 14 to 18 children and two or three teachers. Almost all teachers (96%) were female; they were about equally divided between Black and White ethnic groups, with few from other backgrounds. Three-quarters of the teachers were between 22 and 45 years of age. Although most of the teachers had had considerable general education (61% had a bachelor's degree or higher), few had any specific training in early childhood education (35% had none at all). Moreover, in 1968-1969, most of the teachers had had little paid experience with disadvantaged children (75% had none) and little special training for their Head Start positions (52% had none).

Classroom orientation characteristics. Teachers' self-reported descriptions of their classes reflect an eclectic, child-centered, "whole child" philosophy toward early education. Table 4 indicates the percentages of children in 1968-1969 (totalling 1,648 children) who received either weak, moderate or intense levels of a particular program focus: child-centeredness, task-centeredness and parent-centeredness. As shown in Table 4, most children attended Head Start classes that were moderately to intensely child-oriented, and only weakly task- or parent-centered.

TABLE 4

Level of Program Focus	Weak	Moderate	Intense
Child-centered	19%	34%	47%
Task-centered	75%	19%	6%
Parent-oriented	96%	4%	0%

Teachers also reported on their specific program goals and emphases: language (defined as provision of organized instruction sessions); child socialization (defined by cooperative play, trust toward adults, participation in small groups); independence and self-care (defined by speaking-up for one's own rights and tending to one's own physical needs). Although classes varied greatly in program emphases, relatively few children attended classes having very strong, well-articulated program emphases as indicated by the relative percentages of children, as shown in Table 5.

Trained observers' ratings of ongoing classroom activities largely substantiate the teachers' self-reports. The most frequently observed activities focused more on global aspects of child development than on specific behaviors or skills. These activities, from most to least frequent, included the following: motivational development, self-esteem development, academic skill acquisition, concept development, group and social skills acquisition and last, language development.

According to observers, classrooms also differed substantially in a dimension that was related to implementing special program emphases -- the availability and adequacy of various types of equipment (see Table 6).

As shown in Table 6, most children attended classes which had the equipment (large muscle to encourage gross motor activities such as play, but had less equipment appropriate for specialized activities in cognitive development or learning.

Summary. Teachers' self-reports and trained observers' ratings were in agreement with respect to the general personal-social development orientation that characterized most Head Start classes. Structured cognitive development activities and language instruction were rarely observed, extensive equipment for these activities were rarely present, and strong, well articulated program emphases on cognitive or linguistic attachment were seldom reported. This picture is consistent with the view of Head Start as a broadly based child development program; it does not reveal an intensive compensatory education program aimed at cognitive deficits.

There is always the danger that even observational data, carefully obtained and reported, may fail to capture the essence of the classroom experience or the more subtle but critical variables that define a program orientation. Anecdotal records by the observers suggest that this may sometimes be the case. For example, compare these two reports:

> On the whole, Teacher X had a disciplined and quiet class. She had a rather formal view of education, a mini-grammar school atmosphere ... She emphasized children's ability to communicate, giving them tasks such as to describe to the class what they had done

TABLE 5

Level of Program Emphasis	Lower	Middle	Upper
Language instruction	59%	36%	5%
Child socialization	49%	38%	13%
Independence and self-care	88%	10%	2%

TABLE 6

Availability and Adequacy of Equipment	Low	Middle	High
Cognitive materials	23%	59%	18%
Large muscle	3%	56%	41%
Sensory-motor	17%	64%	19%

over the weekend ... She seems to feel that her job as a teacher ends after she has taught the children their A's, B's, and C's in a group, and her job as a supervisor begins when they are creating things or working on their own.

The children in Miss F's Head Start class seemed a lot happier than children in other centers which I observed. With the exception of about two children, the teacher-child relationship was excellent. The children loved their teacher ... The best thing about this class seems to be the mutual love between all the children. All the children loved each other and they all interacted with each other.

These anecdotes illustrate, among other things, the degree of variation among Head Start classes on several unmeasured interpersonal dimensions, and support the claim that Head Start is many programs, not a single program. There is certainly no evidence of a single educational "treatment" (that is given to all children) in the sense that Sesame Street is a "treatment." The anecdotal data further point out that the nuances of teacher style and classroom atmosphere were barely glimpsed in either the observer forms or teacher self-ratings. Despite the subtle, and sometimes not so subtle, variation in teacher style, it is fair to say that most sampled Head Start classes in 1968-1969 were oriented in a nonspecific way to general personal-social development. It is the effectiveness of this whole-child approach that is reflected in the performance gains, and variations within this approach which are associated with differential gains for some children.

POST-TEST PERFORMANCE GAINS

Did Children Gain From Their Head Start Experience?

The question addressed in this section asks whether Head Start children, taken as a whole, benefitted from their Head Start experience. The answer is yes, in many areas. Those areas in which children showed consistent and statistically reliable gains are summarized in Table 7; those areas in which children's participation in Head Start did not make a difference, i.e., bring about a change in performance, are summarized in Table 8.

Areas showing significant change. Head Start children showed the greatest change or most substantial improvement in areas related to cognitive and social functioning and specific skills -- achievement motivation, social adjustment, school readiness and academic achieve-

TABLE 7

Pre-Post Changes in Performance for 1968-69 Sample

Measures Showing Substantial Change*

Measure	Number Children Tested	Performance Score	Distribution of Children		Mean Score		Direction of Change
			Pre	Post	Pre	Post	
Stanford-Binet IQ (Cognitive Ability)	(1466)	101+	22%	31%	89.5	94.1	+
		90 to 100	27%	30%			
		89 or lower	51%	39%			
Preschool Inventory-Standard Score (School Readiness)	(1232)	121+	31%	54%	109.9	120.3	+
		101 to 120	37%	34%			
		100 or lower	32%	12%			
Preschool Inventory-Sub-Tests (Academic Achievement)	(1233)	personal social	11%	24%	10.7	12.9	+
	(1231)	vocabulary	8%	22%	5.3	7.3	+
	(1229)	numerical	4%	13%	6.0	7.7	+
	(1228)	sensory	19%	50%	11.6	14.8	+
Animal House-Standard Score (Ability to Learn)	(1183)	12+	13%	19%	8.5	9.2	+
		9-11	35%	40%			
		8 or less	63%	40%			
Hertzig-Birch-Verbal Score (Cognitive Style)	(1272)	high	35%	52%	56.4	61.2	+
		middle	50%	41%			
		low	15%	7%			
Factors Affecting Test Performance (Social Adjustment)		very high	7%	66%			
		high	66%	23%			
		moderate	27%	11%			
Gumpgookies-Standard Score (Achievement Motivation)	(982)	high	6%	70%	34.4	41.6	+
		average	21%	19%			
		low	73%	11%			

TABLE 7 (Cont.)

Measure	Number Children Tested	Performance Score	Distribution of Children		Mean Score		Direction of Change
			Pre	Post	Pre	Post	
Parent Interview- (Attitudes toward Education)	(879)	very favorable	34%	54%	125.4	142.1	+
		favorable	41%	33%			
		moderately favorable or less	25%	13%			

*Taken from SDC final report, 1972; Achievement Motivation (Gumpgookies) data from RTI final report, 1972.

TABLE 8

Pre-Post Changes in Performance for 1968-69 Sample

Measures Showing Little Change*

Measure	Number Children Tested	Performance Score	Distribution of Children		Mean Score	
			Pre	Post	Pre	Post
Hertzig-Birch-Work Score (Cognitive Style)	(1272)	96 to 100	26%	24%		
		91 to 95	34%	36%		
		90 and below	40	40	90.3	90.7
Hertzig-Birch-Spontaneous Responses Score (Cognitive Style)	(1272)	high	29%	27%		
		moderate	35%	42%		
		low	36%	31%	111.6	111.7
Sociometric Play-Isolate Score (Sociability)	(1275)	has many friends	80%	79%		
		has some friends	20%	21%	94.1	93.8
Parent Interview (Attitude toward Head Start)	(876)	100% favorable	89%	92%	194.8	195.7
Parent Interview (Feelings of Personal Power)	(887)	high	28%	27%		
		moderate	47%	49%		
		low	25%	24%	123.6	122.8
Parent Interview (Community Involvement)	(887)	moderate or higher	12%	14%		
		low	44%	38%		
		very low	44%	48%	9.4	9.3
Parent Interview (Parent Alienation-high is more alienated)	(881)	19 or higher	32%	35%		
		11-18 (moderate)	62%	58%		
		0-10 (low)	6%	7%	16.6	16.8

*Taken from SDC final report, 1972.

ment. Statistically significant, but less substantial gains were noted in areas related more to basic cognitive abilities -- intelligence or general cognitive ability and learning ability.

The largest gains were made in achievement motivation and social adjustment. At the time of initial testing, only 7% and 6%, respectively, of the children performed at "favorable" levels; in Spring of 1969, 70% and 66%, respectively, performed very favorably on these measures. The performance increment or gain for children receiving very favorable scores was a substantial 64% and 61%.

The next largest gains were reported for skills associated with scholastic readiness: a 17% gain for children obtaining high scores on verbal skills and a 23% gain for children obtaining high scores in school readiness. At the time of entry into Head Start, only 31% of the children demonstrated skills usually predictive of successful school performance; in Spring, 1969, 54% demonstrated such skills.

By contrast, gains in basic cognitive and learning abilities, although statistically reliable, were smaller in absolute magnitude. The performance increment, again in terms of gain in the percent of children achieving very favorable scores at Spring re-testing, was only 9% on the Binet IQ test and 6% on the test of ability to learn. It is worthwhile to mention that the absolute IQ gain of 5.6 points, and the final mean IQ level of 94.1, are similar to those reported in other studies of the effects of traditional preschool experiences on disadvantaged children (Bissell, 1971; Stanford Research Institute, 1971; Stearns, 1971). And, existing evidence supports the argument that gains of this magnitude are best accounted for by motivational rather than cognitive changes associated with these experiences (Zigler and Butterfield, 1968). Although the final level of performance reported by this national evaluation is still below the national average (about 6 points below) assessments of individual Head Start programs have reported larger gains and final levels closer to the national average (Ryan, 1974).

It seems likely that the changes on cognitive and social measures presented here are "psychologically meaningful" in the sense of reflecting real phenomena as distinct from testing and regressing artifacts. To illustrate this point, compare the very large gains in initially high levels of achievement motivation (from 6% to 70%) with the less substantial gains in initially high scores on tests of school readiness (from 13% to 54%) or tests of the ability to learn (from 13% to 19%). Although ample room for improvement was available on these measures, the magnitude of gains varied considerably from one to another suggesting that the gains do not simply reflect a tendency toward greater improvement where greater room for improvement exists.

Areas showing little or no change. Table 8 summarizes the results for those measures which did not change during the Head Start program.

Many of these no-change measures tapped parent attitudes. It appears that while parents' attitudes toward education improved from "favorable" to "very favorable", parents' attitudes vis-a-vis society did not change: no significant improvements were found in parents' feelings of personal power (predominantly externally and not internally controlled), of reported community involvement (predominantly low or very low), or of alienation (predominantly moderate or high). Thus, it does not appear that Head Start affected how parents felt about their own chances or place in society, although there was ample room for improvement on all measures of these attitudes. These findings should not be interpreted to mean that Head Start has no effect on parents' lives; many anecdotal reports of parents tell of dramatic changes from fearful, isolated people to dynamic community leaders and competent self-assured individuals. On the other hand, the anecdotal reports may be somewhat atypical. Another caveat to bear in mind is that the durability of changes in parent attitudes after termination of the Head Start experience is unknown at this time.

Also depicted in Table 8 are several measures of children which did not show reliable change. Two of these were virtually already at "ceiling level" (i.e., did not leave much room for improvement): measures of the child's willingness to work on a test such as the Binet and peer sociability (non-social isolation), suggesting that these are not among entry-level problems of Head Start children.

Summary. Overall, then, children began Head Start low in achievement, in scholastic readiness and general learning ability. At the end of six to eight months, in Spring 1969, they were highly motivated, very well-adjusted to the school situation, had acquired satisfactory levels of school-readiness skills, and showed modest gains in general learning ability. Their parents, moreover, had very favorable attitudes toward education (20% increment in parents with very favorable attitudes toward education).

Do these gains reflect the Head Start program emphasis on personal-social development reported by teachers and observers? Or do they indicate that basic cognitive abilities are much more difficult to influence than motivation? The clearest way to resolve these questions would require carefully controlled experimental studies in which initial levels of entry characteristics were varied systematically with contrasting program emphases. Then, if a particular program emphasis produced a particular pattern of gains -- e.g., a personal-social program emphasis produced the largest personal-social gains, but smaller cognitive gains, and a contrasting program emphasis produced an opposite pattern of gains it could be reasoned that program emphasis accounts for differences in the magnitude and type of gains made. On the other hand, if the same pattern of gains resulted from contrasting program emphases-- e.g., for all programs the largest gains would be reported for motivation; next largest for specific skills; smallest for basic cognitive abilities--it could be reasoned that program emphasis is irrelevant to the type, if

not magnitude, of gains made. It is unlikely that any experimental study would yield such clear cut findings, and permit drawing such unambiguous conclusions about a complex issue of this sort. It is even less likely that the loosely controlled national evaluation studies summarized in this report will resolve completely the issue of whether "curriculum" or program type makes a difference to the gains obtained in early intervention programs. However, the results of these evaluation studies should shed considerable light on this issue.

Did Some Children Gain More Than Others?

Overall, the answer is yes. On the Stanford-Binet IQ test, for example, changes ranged from losses of more than seven points for 9.5% of the children to sizeable gains of more than 18 points for 9.2% of the children. The remainder of the data analyses were addressed to identifying the factors associated with these differences. Three kinds of influencing factors were examined and will be discussed in this report:

--background characteristics of the children and their families
--characteristics of the teachers and programs
--interactions or unique combinations of child/family and teacher/program characteristics.

Statistically significant and reliable ($p < .01$) changes associated with these factors will be presented and discussed. The main data base will continue to be the 1968-69 national evaluation study and where appropriate and possible, reliable relationships from the 1967-68 and 1966-67 data will be mentioned.

Which Child and Family Characteristics Are Associated with the Greatest Gains In Performance?

The purpose of this section is to explore the issue of whether different types of children make different kinds of gains in the Head Start program. If children do vary on a particular characteristic, for example age, vary systematically in their performance gains (i.e., gains vary in relation to differences in age), it can be reasoned that age of Head Start children makes a difference to the gains they will make. On the other hand, if the differences in children's age are not systematically related to performance gains, an argument can be made that age differences may not be an important determinant of the gains made in the Head Start program.

The characteristics of children and families measured in these studies are listed below with the relationships that were found between these characteristics and performance gains.

.Age. Younger children (under 60 months of age) gained more in the area of social adjustment; older children gained more in specific

skill areas (e.g., numerical). Younger children also made greater gains on the Stanford-Binet test of general cognitive ability in both 1966-67 and 1967-68 data, but curiously, this pattern was not replicated in the 1968-69 data.

.Sex. Whether the child was male or female was unrelated to performance gains on all measures. Not only were sex differences unrelated to gains, but recall as well the lack of significant relationships with initial performance levels. These findings fail to support the assertion that the all-female teaching staff is associated with the unfavorable program experience often reported for boys. Several interesting and unanswered questions are suggested by this intriguing no-difference finding and perhaps warrant further probing in subsequent statistical analyses. For instance, did the male volunteers (whose performance was not analyzed separately) contribute to producing equal benefits for boys and girls? Were teacher aides, many of whom come from the community, responsible for the equality of boys' and girls' experiences?

.Ethnicity.[12/] The relationships between ethnicity and gains were, on the whole, different for different measures, presenting no consistent picture or pattern. For example, on the Stanford-Binet, Black children made the greatest gains and the Mexican-American children the smallest; however, no significant differences between ethnic groups on Stanford-Binet gains were reported in 1967-68. And, on measures of school readiness, Polynesian children made the greatest gains. These findings should be interpreted with caution, however, since the Mexican-American and Polynesian samples were small (139 and 83 children respectively) and the data are confounded -- Mexican-American and Polynesian children differed from their Black and White Head Start counterparts in age as well as ethnicity.

.Initial IQ.[13/] Pretest scores on the Stanford-Binet were related to gains on several measures. In both 1968-69 and 1967-68, children with initially low IQ scores (less than 85) showed the greatest gains on the Stanford-Binet. It is important to note that this relationship was linear, with medium entry IQ's (85-95) associated with moderate gains and high entry IQ's (over 95) with small gains. The same linear relationship was also found for gains on measures of school readiness (1968-69).

12/ The relationships between ethnicity and gains were analyzed only by Research Triangle Institute.

13/ Findings concerning the relatioships between initial IQ and gain scores are based on analyses performed by System Development Corporation.

The strong relationship between initial IQ and IQ gain scores can be partially accounted for by the measurement artifact of "regression-toward-the-mean": extremely low scores may be spurious due to fatigue, distraction, etc. and, on subsequent re-testing, individuals who scored extremely low are likely to score higher, thus raising the mean of the group. However, as concluded in the final report of these evaluations (SDC, 1972), the differential gains made by the low IQ group seem to reflect more than regression artifacts because: (a) the low initial IQ group had the lowest scores on all pretests except social isolation, indicating that placement of a child in this group was not due to random error of measurement on one pretest measure alone, and (b) the low initial IQ group gained most on nearly all performance measures. The strength of this argument is diminished by the fact that the 1967-1968 data do not show this extensive and consistent pattern. The influence of initial IQ is best interpreted, therefore, partly as a result of statistical regression and partly as a result of genuine differential gain.

.Prior Head Start Experience. Children with no previous Head Start experience and children with lengthy previous Head Start experience showed greater gains in preschool academic achievement and ability to learn a new task than did children with short previous exposure to Head Start. This finding is difficult if not impossible to interpret; it may reflect the difference between summer and full-year Head Start, or it may be a result of factors confounded with previous Head Start attendance.

.Urbanicity. On a variety of measures, especially of cognitive functioning, children from non-urban communities showed greater gains than urban children. In 1967-68, these differential gains occurred only in the social domain. However, non-urban children also differed from urban children in their initial scores on these measures. Thus, although these results are tantalizing, no unambiguous conclusions can be drawn about the effects of urbanicity, since these data are confounded.

.Region. The relationships between geographic region and gains were inconsistent between 1967-68 and 1968-69, probably reflecting experimentally induced changes in the Head Start program in the South (1968-69). In 1967, non-Southern children showed larger gains in both cognitive and social-emotional development than did their Southern counterparts (the difference in cognitive gains was only marginally significant). However, in 1968, after many Southern Head Start centers were converted to structured academic programs, Southern children showed greater gains than non-Southern children on measures of verbal responsiveness, specific academic skills (numerical), and cognitive functioning (again marginal in the cognitive area). It seems likely that region per se is not an important determinant of gains, but that the program variables (e.g., academic orientation) and child characteristics (e.g., age) that are confounded with geographic region in this study may indeed be related to gains in a psychologically meaningful way.

.Family Socio-economic Status. Relative socio-economic status within Head Start did not vary greatly and perhaps for this reason few significant relationships between socio-economic status and gains were reported. Nonetheless, there was a consistent (but not statistically significant) tendency for the most "deprived" children to show the greatest gains. A single isolated exception was the significant positive relationship between the mother's educational level and the child's ability to learn a new task. This finding was not replicated in 1967-68, however, suggesting that within the Head Start population, relative socio-economic status is not a determinant of gains.

.Other Family Variables. Unrelated to performance changes made during the Head Start program were the host of family background characteristics, which, by comparison, were strongly related to initial performance levels. These characteristics included: family structure, parental employment, family mobility, maternal aspirations and expectations for the child's education, parent pessimism, accessibility of adults in the home, frequency of reading to the child, parent mode of physical control.

Summary. Although some child and family background characteristics are meaningfully related to performance gains, many more appear unrelated or only inconsistently related to gains. Child's sex and most family characteristics including socio-economic status were unrelated to gains; child's previous Head Start experience, ethnicity, urbanicity and geographic region were only inconsistently related to gains and were confounded with other variables such as program orientation, child's age and prescores. The strongest and most consistent relationships emerged between child's age and pretest IQ and gains on measures of social and cognitive development. However, the very strong relationships between pre-test IQ and gains are tempered by testing artifacts. Two important conclusions can be drawn from these findings: first, that the family characteristics measured in these studies are not important determinants of the gains the child makes during the Head Start program although they are very critical determinants of the child's performance level at the time of entry into Head Start; and second, that, with the exception of age and initial IQ, child background characteristics as measured in these studies also do not appear to be important determinants of gains made during the Head Start program. The relationship of other kinds of characteristics (teacher/program) and combinations of characteristics (child/family and teacher/program) to gains scores will be explored in subsequent sections.

Which Teacher and Program Characteristics Are Associated with the Greatest Gains in Performance?

This question will be approached in two ways: first, in terms of specific relationships between various program and teacher characteristics and children's gains; and second, in terms of the program inputs

associated with more broadly defined child outcome areas. The purpose of organizing the findings this way is to ensure that children's gains are viewed from the two perspectives of program inputs and child development outcomes.

How teacher/program characteristics are related to gains. In this section the following categories of teacher and program characteristics will be discussed:

--teachers' behaviors in the classroom
--classroom activities
--classroom materials
--aspects of classroom atmosphere.

1. Teacher's Background

Teacher's background characteristics were assessed by means of a structured teacher interview. The most significant background characteristics in terms of implications for early childhood education, were the teacher's education and prior experience with disadvantaged children.

Teacher's age. Children whose teachers were older (over 28 years of age) made greater cognitive gains than did children in classes with younger teachers.

Teacher's education and experience. Children whose teacher's level of general education and extent of her prior experience with disadvantaged children were relatively high performed less well on measures of preschool achievement and social adjustment than children having teachers who appeared to be less well qualified. This puzzling finding requires some further explanation and in so doing, will draw upon additional data.

In 1967-68, teacher's education and experience were also negatively related to child outcomes, but in the domain of social interaction with peers. Separate analyses for Southern and non-Southern children revealed, however, that the effects of teacher's experience (but not education) were regional. In other words, the negative relationship between teacher's prior experience and cognitive performance held only for non-Southern children, suggesting possible confounding with other factors. Moreover, there is some indication (from the RTI report, 1968-69) that specific training in early education was actually positively related to gains in school readiness and learning ability, but these effects were, unfortunately, weak and small.

Certification requirements for early childhood teachers have been among the most controversial policy areas. The negative relationship between the years of general education and benefits to children

seems to support the view that paraprofessionals can relate better to Head Start youngsters. However, proponents of this view should also take into account the pervasive lack of prior paid experience with disadvantaged children among Head Start teachers during 1966-69. It seems likely that not the years of general education per se, but the lack of relevant prior experience coupled with it, may account for this curious negative relationship between teacher's qualifications and children's benefits. The data are most consistent with the position held by many early childhood educators that education and experience in early childhood education rather than general education alone, are essential preparation for preschool teachers. However, too few teachers in 1966-69 had degrees or training in early childhood education for the analyses to be more than suggestive. Future analyses of the 1971-72 Planned Variation data and evaluative research on the new "credentialling-by-accomplishment" Child Development Associate Program will be valuable in resolving the question of what training is associated with teacher competence and child benefits.

2. Teacher's Approach and Program Emphasis

Differences among teachers' approaches and program emphases, as reported by the teachers themselves, were related to a variety of outcomes. The relationships are summarized below.

Method of control. Children having teachers who used non-physical control gained more on measures of cognitive abilities and school readiness than children with teachers who used physical control to maintain discipline and order in the classroom.

Teacher's style - cognitive aspects. Children whose teachers stressed providing the children with quality cognitive inputs were not differentially affected; however, the parents of these children showed an increase in their feelings of alienation.

Emphasis on independence and self-care. Children whose teachers reported high emphasis on goals of independent and self-care showed greater gains in cognitive abilities, school readiness, and social adjustment than children whose teachers reported a lower emphasis on these goals.

Emphasis on socialization. Children whose teachers emphasized child socialization goals showed better social adjustment in the testing situation than children whose teachers did not emphasize these goals.

Emphasis on language. Children whose teachers emphasized language development gained less on measures of cognitive abilities and school readiness than children whose teachers did not emphasize language development. This finding is problematic unless viewed in the context

of the classroom observation data. Actual observations revealed that high use of language materials was associated with greater cognitive gains, as would be expected. It appears there is some discrepancy between teacher's self reported program emphasis on language and actual observations of ongoing program activities.

3. Teacher's Behavior in the Classroom

Several characteristids of the teacher's behavior in the classroom were differentially associated with children's gains. These characteristics are described in detail in Appendix A as the teacher factors derived from the 1968-69 data collected with the observational instrument (OSCI - Observation of Substantive Curricular Input). Teachers' behavior was, on the whole, significantly related to gains in many areas, as indicated by the relationships summarized below.

Social-emotional interactions with children. Children whose teachers engaged in more social-emotional interactions did not perform as well on measures of the ability to learn a new task as did children whose teachers exhibited less of this behavior.

Structured lessons. Children whose teachers structured lessons to a greater extent, showed greater gains on measures of learning ability and school readiness.

Art activities. Children whose teachers provided more art-related activities showed greater gains in school readiness.

Creative small-group instruction. Children whose teachers used middle and high levels of creative small-group instruction activities (e.g., visual motor activities) showed greater gains in cognitive ability, school readiness and the ability to learn a new task than children whose teachers provided only low levels of such activities.

Rote learning. Children whose teachers placed a relatively low emphasis on rote (receptive) learning showed greater gains on measures of learning ability.

4. Classroom Activities

Characteristics describing classroom activities were also derived from the OSCI observational instrument. Compared with teachers' behavior, however, classroom activities (see Appendix A, 1968-69, OSCI class factors) appear to be less closely related to children's gains. In fact, several factors (the presence of group activities in a social context; social-emotional interactions across contexts; and small group creative activities) were altogether unrelated to gains on any outcome measure. Classroom activities that were differentially associated with gains are summarized briefly below.

Structured lessons. Children in relatively "structured" classes (e.g., programmed materials and control by head teacher) showed greater gains in school readiness and learning ability.

Verbal communication. Children in classes characterized by fairly extensive verbal communication across contexts gained less on measures of learning ability and achievement motivation.

Language and discrimination learning. Children in classes characterized by frequent occurrence of language and visual discrimination activities and use of learning materials gained more in general cognitive ability but less in learning ability.

Parent participation. Children in classes where parents participated in the program (e.g., as defined by the number of visits to the Head Start center) showed greater gains in the area of social-emotional adjustment. The relationship was weak, however, and represents only a very token measure of parent involvement.

5. Classroom Equipment and Materials

The physical attributes of the classroom were yet another aspect of Head Start programs measured in the 1968-69 and 1967-68 studies. Classroom equipment and materials were related to gains, but the data suggest two possible interpretations: first, that classroom equipment and materials are incidental indices of the physical quality of the program; or second, that the equipment and materials themselves are psychologically related to the gains reported. The first interpretation is most fitting in the case of large muscle equipment, described below.

Large muscle equipment. Children in classrooms with access to large muscle equipment (e.g., jungle-gyms) gained more in cognitive ability and school readiness than children in classes where such equipment was less accessible. However, observed high frequency use of this equipment was associated with smaller gains in cognitive ability, fewer social interactions among children, and less favorable parent attitudes about the perceived benefits of Head Start. While the presence of this equipment is associated with effective programs, excessive use of the equipment may divert the children from activities more likely to enhance their cognitive development.

Small muscle equipment. Children in classrooms where small muscle equipment was available and extensively used showed increases in the frequency of social interactions initiated with peers (1967-68 data).

Cognitive materials. Children in classrooms where cognitive materials were available and extensively used, showed greater improvement in their task orientation in the testing situation.

Dramatic materials. In classrooms where dramatic materials were used extensively (and considerable verbal interaction probably occurred), children showed comparatively greater increases in social interactions and parents revealed more favorable perceptions of the benefits of Head Start.

6. Classroom Atmosphere[14/]

Classroom atmosphere includes characteristics which reflect more qualitative and experiential aspects of the classroom environment, and which have little to do with programmatic inputs. Nonetheless, such characteristics are related to differences in gains.

Student and teacher turnover rates. Children in classes where the student turnover rate was relatively low showed greater gains in cognitive ability, social interactions with peers, and social adjustment than did children in classes with high transiency. On the other hand, children who had experienced a high degree of teacher continuity also showed greater improvement in their social adjustment. This pattern suggests that children's social adjustment in stressful situations (specifically, testing) is facilitated by a changing peer context, but within a relatively stable teacher-child context.

Length of exposure to Head Start. While there were no substantial relationships between pre-post interval length and gains, longer intervals (defined as 27 or more weeks) were associated with greater cognitive gains and improvements in social adjustment and behavior problems. These findings suggest that program effects are cumulative, at least within the time span of the full-year Head Start program.

Which Program Inputs Are Associated with Specific Outcomes?

The numerous and often complex relationships between teacher/program characteristics and child outcomes presented in the previous section of this report will be restated below in more general terms by clustering the program inputs around areas of groupings of child development outcomes.

14/ Data pertaining to student and teacher turnover rates are taken from 1967-68 and 1966-67 final reports, System Development Corporation; data pertaining to length of exposure to Head Start are taken from the 1967-68 and 1968-69 final reports, Research Triangle Institute.

1. Preschool Achievement

In general, children gained most in classes where:

- .the teacher was older
- .the teacher had less formal education and experience
- .the teacher did not use physical control
- .there was high or moderate emphasis on independence and self-care
- .there was high emphasis on structured lessons
- .there was moderate emphasis on small group activities
- .there was ample large muscle equipment available but not excessively used
- .there was high emphasis on art activities

2. General Cognitive Ability

In general, children gained most in classes where:

- .the teacher was older
- .there was moderate emphasis on small group activities
- .there was low emphasis on rote learning
- .the teacher did not use physical control
- .there was high emphasis on independence and self-care
- .there was ample large muscle equipment which was not excessively used
- .there was moderate use of cognitive materials
- .despite the teacher's self-reported low emphasis on language, language materials and activities were frequently observed

The value of moderate and high levels of cognitive structure and activities for cognitive gains was replicated in the 1967-68 data as well.

3. Learning Ability

In general, children gained most in classes where:

- .the teacher was older
- .there was a low emphasis on teacher's social-emotional interaction
- .there was high emphasis on structured lessons
- .there was low emphasis on rote learning
- .there was low emphasis on verbal communication
- .there was moderate emphasis on small group activities
- .there was low emphasis on language and discrimination learning tasks

4. Social Adjustment

Social adjustment, as an area of outcomes, encompasses adjustment to the testing situation, reduction in the number of behavior problems, and improvement in motivation and self-esteem. In general, children gained most in classes where:

- .the teacher had little formal education or previous experience
- .there was high emphasis on independence and self-care
- .there was high emphasis on child socialization
- .there was a high level of teacher continuity and a low pupil/teacher ratio

The clusters described above show how teacher and program inputs, including teacher's educational background, approach and classroom behavior, were related to children's gains in all outcome areas. Classroom activities and materials, independent of the teacher's input, were on the whole less extensively related to child outcomes, particularly in the area of social adjustment. Cognitive gains were influenced more by cognitive-oriented program emphases and classroom activities; social gains were influenced more by social-oriented program emphases and by qualitative aspects of classroom atmosphere.

Do Different Teacher and Program Approaches Influence Different Children in Different Ways?

Experts in the field of early childhood education disagree on the definition of a good preschool program. According to Weikart (1971), and Weber (1971), curriculum is less critical to child outcomes than the non-substantive characteristics of the program such as administration, planning, and supervision. Some (Bissell, 1970; Kohlberg and Mayer, 1971) have identified global substantive program attributes (e.g., program structure; program emphasis on facilitating stages of cognitive development) that seem effective in a broad sense, in facilitating child development. Still others (e.g., Miller, 1971) have found that different programs produce different effects on different children -- interactions -- a finding that supports the position held by most early childhood educators that programs should be "matched" to child characteristics.

Analyses of the Head Start national evaluation data do reveal interactions between program and child characteristics.[15] A program or teaching approach associated with superior performance for one group of children may not necessarily be associated with superior performance for another group; and in some instances, program inputs that are positively related to gains for one group may be negatively related to gains on

15/ The interactions summarized in this section of the paper were reported by SDC only. SDC analyzed the data using an approach

other measures or gains for another group of children. To restate the point, interactions are oftentimes complex inter-relationships between two kinds of characteristics, in this case program and child characteristics, and outcomes.

Teacher/program approaches associated with gains for different groups of children. In this section interactions between program and child characteristics will be presented to highlight the particular program approaches or teacher characteristics which were found to have positive effects on certain subgroups of children within the Head Start population tested.

.Teacher's use of non-punitive methods of control or restraint from physical punishment, was most consistently related to gains in cognitive abilities and preschool achievement for:

--children with initially high IQ's (over 95)
--relatively younger children (under 5 years of age)
--urban children

.Moderate to high levels of program emphasis on goals of independence and self-care were positively associated with gains in cognitive abilities, achievement and social adjustment for:

--children with initially high IQ's
--younger children
--urban children

.Moderate levels of program emphasis on child socialization goals were associated with gains in achievement for:

--children with initially high IQ's
--younger children

High levels of this program emphasis also benefited children with initially low IQ's (less than 85) on their social adjustment in the testing situation.

15/ (cont.) especially sensitive to detecting interactions -- each combination of teacher/program and child variable (which included pretest IQ, age, urbanicity, region, and pretest-posttest interval length) was considered separately. The statistical approach used by RTI, on the other hand, was not especially sensitive to interactions and their report concludes that gains were primarily accounted for by variations in program/teacher approaches, not by combinations of program/teacher approaches and child characteristics. Nonetheless, the direction of the relationships between programs and gains reported by RTI is consistent with the interactions summarized in this report.

.High levels of program emphasis on language development were found to have mixed and even reversed effects: a strong program emphasis was positively associated with gains in achievement and cognitive abilities for children with initially low and mid-level IQ's, but negatively associated with cognitive gains for children with initially high IQ's.

For younger children, the strong language development emphasis negatively related to gains in both cognitive abilities and cognitive styles. But, conversely, a strong language development emphasis was associated with achievement gains and cognitive style gains for older children.

.The quality of teacher's cognitive input also displayed reversal effects. Moderate quality cognitive input, as reported by teachers, was associated with the greatest gains in cognitive abilities for children with high pretest IQ's. However, high quality cognitive input was also associated with the intensification of feelings of alienation among parents whose children were more disadvantaged: those with initially low IQ's, and those who were older at the beginning of the Head Start program.

.Extensive use of dramatic materials (in 1967-68), which are distinguished by a large verbal component, was related to gains on indices of overall verbal behavior for:

--children with initially low and mid-level IQ's
--both younger and older children
--urban children

Interestingly, this program activity was also associated with positive changes in measures of parents' perceived effects of Head Start among parents whose:

--children had initially high IQ's
--children were younger
--children were non-urban

.The availability and adequacy of large muscle equipment was related positively to gains in cognitive abilities for children with initially low and mid-level IQ's; to achievement gains for older children; and to improvements in cognitive style for non-urban children.

However, the actual use of large muscle equipment was negatively associated with gains in cognitive abilities for these same groups of children and negatively related to parents' perceived effects of Head Start for all subgroups of children. As in an earlier section of this paper which summarized overall program effects, these two superficially similar variables display

opposite effects, particularly on the more disadvantaged Head Start children. Recall that availability can be construed merely as an index of program quality whereas actual use of large muscle equipment indicates the nature of children's experience in the program.

.A high degree of teacher continuity (in 1967-68) was associated with improvements in social adjustment in the testing situation for:

--children with initially low IQ's
--younger children
--urban children

.And, in 1967-68, a smaller pupil/teacher ratio was related to gains in both social adjustment to testing conditions and sociability among peers for:

--children with both initially high and low IQ's
--younger children
--urban children

Ratios ranged from a low of 3:1 to a high of 18:1 with over two-thirds of the children in classes having ratios between 5:1 and 9:1.

In these same groups of children, class transiency, or pupil turnover rate, was negatively related to gains in both cognitive abilities and sociability among peers suggesting that for some subgroups of children a stable peer environment appears to be more facilitative.

.Negative relationships occurred between the level of the teacher's general educational preparation and gains in preschool achievement and social adjustment (in both 1967-68 and 1968-69 data) for:

--initially low IQ children
--older children
--urban children

.Although teacher's paid experience with disadvantaged children was unrelated to gains in cognitive abilities, it was negatively related to gains in achievement, social adjustment and sociability among peers for:

--children with initially low and mid level IQ's
--younger children
--urban children

It is noteworthy that parental feelings of power were negatively related to this variable and in all subgroups of children analyzed. As discussed earlier, the results associated with these variables are problematic, partially because the variables are not measures of specific early childhood education or specific teacher competencies.

What Kinds of Children Benefited Most from Different Teacher and Program Approaches?

In this section program-child interactions will be presented to highlight the particular groups of children for whom certain clusters of program and teacher characteristics are most effective. Consistent with all preceeding analyses the subgroups of children are differentiated according to initially high or low IQ, younger or older within the Head Start population, and urban or non-urban location.

.For children with initially low IQ's (below 85) cognitive and social gains were greatest in programs where:

--child socialization goals and language development activities were strongly emphasized
--the teacher used non-physical means of control
-large muscle equipment was available and adequate but not excessively used
--classroom atmosphere reflected teacher continuity, a low pupil/teacher ratio and low pupil turnover rate
--teachers had less formal general education

.For children with initially high IQ's (over 95) cognitive and social gains were greatest in programs where:

--the teacher reported a moderate quality of cognitive input to children
--the teacher did not rely on physical means of control or discipline
--a high emphasis was placed on goals of independence and self-care, and a moderate emphasis was placed on socialization goals
--the classroom atmosphere was characterized by a small pupil/teacher ratio and low pupil turnover rate

.For younger children (under five years of age) cognitive and social gains were greatest in programs where:

--the teacher relied on non-punitive and non-physical methods of control
--a moderate to high emphasis was placed on child independence, self-care and socialization goals

--dramatic materials were frequently used
--the teacher reported a low program emphasis on language development
--classroom atmosphere reflected high teacher continuity, a low pupil/teacher ratio and low pupil turnover rate
--the teacher had achieved a lower level of formal general education

.For older children (over five years of age) cognitive and social gains were greatest in programs where:

--a strong emphasis was placed on language development
--dramatic materials were extensively used
--adequate large muscle equipment was available but not used excessively
--the teacher had achieved a lower level of formal general education

.For urban children cognitive and social gains were greatest in programs where:

--the teacher used non-punitive, non-physical methods of control
--a moderate to strong emphasis was placed on child independence and self-care
--dramatic materials were extensively used
--the classroom atmosphere reflected a high degree of teacher continuity, a low pupil/teacher ratio, and low pupil turnover rate
--the teacher had achieved a comparatively lower level of formal general education and experience with disadvantaged children

.For non-urban children cognitive gains were greatest in programs where:

--adequate large muscle equipment was available but not excessively used

Significance of the program-child interactions. The picture that emerges from these analyses reflects the complexity of the interrelationships among children, teachers, facilities, and programs. The findings suggest that there may be two constellations of program variables which are differentially effective for different groups of children. 16/

16/ Technically, the relationships listed hold for each separate program element and each child characteristic individually. No direct test was made of the specific combinations of program elements (e.g., programs which had large muscle equipment and strong language emphasis), but there is no reason to believe that clusters of program elements would produce effects different from those of the elements considered individually.

The first cluster includes program elements related to the orderliness and structure of the total environment of the Head Start center. Specifically, these elements include: high quality facilities (as reflected in the availability of expensive large muscle equipment); low pupil/teacher ratios; high teacher continuity; and moderate to strong program emphases (including academic emphases). This program cluster seems to be especially beneficial for children who begin Head Start with a relatively low pretest IQ (below 85), children who are older, and children from both urban and non-urban areas.

The second cluster includes program elements more related to the social and interpersonal climate of the class. These include: teacher's use of non-physical methods of control; moderate, socially-oriented program emphases (on independence, self-care and socialization); a stable classroom atmosphere characterized by high teacher continuity, low pupil/teacher ratios and low pupil turnover rate; and more frequent use of dramatic (role playing) activities. Children who especially benefit from these program/teacher inputs had high pretest IQ's relative to their Head Start counterparts, entered Head Start at an earlier age (under five years) and tended to live in urban areas. 17/

These two clusters of program elements are reminiscent of the two hypothetical clusters of program variables identified by Stodolsky (1971) in a review of the results of one site research and evaluation reports. Stodolsky, however, reports that only the first of the clusters results in significant gains; these findings indicate that both program types are effective, but for different kinds of children.

It should be noted, however, that these two clusters do overlap somewhat. They do not identify two completely distinct and independent program orientations which are beneficial for two different and non-overlapping populations of children. A direct experimental test of the effectiveness of these two program clusters, with random assignment of children varying in IQ, age and urbanicity, is needed before the pattern observed in these data is adopted as a sound educational policy and basis for program prescriptions.

17/ In the national Head Start program, and in these evaluation data, child's age, IQ, urbanicity, and geographic region are confounded. In the South, few school districts have public kindergartens and thus the children in the South are four and five year olds while children from other regions are three and four year olds. IQ and age are negatively correlated in this sample. However, this probably does not reflect a decline in IQ performance with age. Vander Riet (1972) found that within a Southern sample, cross-sectional data showed a grade level decline in IQ while longitudinal data did not. It seems more likely that the frequently reported lower performance of children from the South (even when matched for age, income, and ethnicity) on a variety of achievement measures is responsible for the age/IQ correlation observed repeatedly in National Head Start data.

Notwithstanding these caveats, the similarity of these program clusters to the recommendations of most early childhood educators is striking. Essentially, the findings suggest that young, bright children who live in high resource areas (urban) benefit most from a preschool experience that encourages motivational independence and opportunities to initiate creative activities, while older and less able youngsters who live in less well endowed environments, benefit more from the program structure, classroom materials, and opportunities to develop specific skills -- the difference between activities in a nursery school and those in a kindergarten. These data do not indicate what will benefit the younger and less able child or the older and brighter child since age and ability are confounded. As in the preceding instance, a direct experimental study is needed to provide hard evidence on this issue.

DISCUSSION

Factors Influencing Developmental Changes

Most significant in the Head Start evaluation data collected from 1966-1969 are the documentation of a broad range of developmental gains over a broad range of Head Start programs and the implications concerning which factors influence differences in program effectiveness. Since no control data were obtained, these findings should be construed only as statistical associations and not causal relationships between program inputs and child development outcomes. Nonetheless, these evaluation data do provide considerable information about the relative benefits of different types of Head Start programs for different types of children. From a research perspective, the data contribute most to the generation of increasingly specific hypotheses about early childhood intervention. From a policy perspective, the findings also provide some tentative guidance for early intervention programs and evaluation.

In the design of these evaluation studies, four types of factors were examined for their relative contribution to the developmental changes associated with participation in the full-year Head Start classrooms. These factors were: family background characteristics, child characteristics, program and teacher characteristics, and combinations of program/teacher and child characteristics. The nature of the influence of these factors will be reviewed separately below.

Family background characteristics. Although family background characteristics were shown to be an important determinant of initial levels of performance, their influence on changes in performance once in the program was negligible. The family characteristics most strongly associated with initially higher levels of performance reflect the dynamic aspects of family life -- the quality of child's experience in the family and the nature of the parent-child relationship. These family characteristics included high parental aspirations of expectations for the child, a sense of confidence about life, adult accessibility to the child, and restraint in using physical punishment as a means of disciplining the child.

Within the narrow range of socio-economic status of the families studied, one characteristic was positively associated with changes made in the program -- the mother's level of education. This single finding is important, however, since it is consistent with those reported for most early intervention programs: program effects are smaller (and lost earlier) among the more deprived children (see Ryan, 1974; Bronfenbrenner, 1974). However, it should be mentioned that with the single exception of mother's education these evaluation data showed a more general, though statistically non-significant, tendency for the more deprived children to show the greatest gains. On balance, the

findings with respect to socio-economic status and gains are problematic.

Characteristics of the child. Compared to the family background characteristics discussed above, background characteristics of children (i.e., age, sex, ethnicity, urbanicity and geographic region) were unrelated or inconsistently related to initial differences in performance levels. But unlike family characteristics, some child characteristics were found to be important determinants of performance changes made during the program. Of the attributes measured in these evaluation studies, child's relative age (i.e., below or above five years) and pretest IQ level (below 85; 85-95; or above 95) were significantly associated with gains on several measures of social and cognitive development.

Younger children gained more in social adjustment and general cognitive ability whereas older children gained more in the acquisition of specific skills. These findings are consistent with the view expressed by many child development professionals that what a child gains from a program depends on the one hand, upon his competence level and stage of development and, on the other hand, the nature of the experiences provided to him/her in the program. These data show differing patterns of gains for older and younger children and do not support the conclusion that younger children gain more across-the-board. Bearing in mind that these data sample a fairly restricted age range of children (approximately two and one-half to six years) and are confounded with other variables measured in these studies, they do suggest that relative age of the preschool child is less a determinant of size of gains than a determinant of type of gain made in the program.

The findings with respect to the relationship between child's initial level of IQ and gains are especially problematic because of the presence of testing artifacts, as discussed previously. The findings are in agreement with those commonly reported in other intervention research and show the expected greater IQ and achievement gains for children with the lowest entering IQ's. But one can argue, as Bronfenbrenner (1974) has, that the data are not sufficient to claim that the least competent children gain most from early intervention. As Bronfenbrenner has pointed out, the existing evidence is based on studies which have not adequately controlled for regression artifacts, with the exception of a study by Herzog (see Ryan, 1974) that shows no long term differences between initially high and low IQ groups. Yet, in the 1966-69 evaluation findings, the relationships between initial IQ levels and gains are so strikingly strong and linear that they lend some credence to the position that, despite testing artifacts, less competent (in terms of IQ) children do seem to make greater real gains, especially in cognitive areas.

Teacher and Program characteristics. The major determinants of gains made in the Head Start program include a multiplicity of teacher and program characteristics. Teachers' background, teacher's approaches and program emphases, teachers' behavior in the classroom, classroom activities, equipment, materials, and classroom atmosphere were all

related to child development gains in preschool achievement, general cognitive ability, learning ability and social adjustment.

In general, moderate and high levels of cognitive inputs or emphases yielded greater gains in cognitive ability and preschool achievement. This finding is consistent with those reported by Bronfenbrenner (1974) and Stearns (1971) for other preschool intervention programs, and supports the position that "cognitive oriented" or "cognitive structured" programs produce greater cognitive gains. It should be noted however, that the precise meaning of cognitive structure varies from program to program and study to study, and that in the context of most personal-social oriented Head Start programs, cognitive structure seems to refer to teachers' emphasis on structured lessons, cognitive materials, and small group activities.

It is noteworthy that these findings confirm that program elements or curricula do make a difference to program impact, both in terms of magnitude and type. The influence of program variables is shown in two ways in these data. First, the overall pattern of gains parallels, almost exactly, the pattern of classroom activities that characterized most Head Start programs, i.e., the extensiveness of a particular program emphasis and the magnitude of gains corresponded for achievement motivation, social adjustment, school readiness, academic achievement, and finally, cognitive abilities. Second, more differentiated analyses revealed that even within the predominant personal-social context of most Head Start classes, particular program elements were differentially associated with gains, e.g., more cognitively oriented programs tended to produce larger cognitive gains.

These data also provide a tentative and partial answer to the question posed earlier on the relative contribution of program emphases or the ease of modifying certain areas to differential gains in motivation and cognitive abilities. The data show that program emphases do account for the greater gains in motivation and achievement as compared to gains in cognitive abilities. However, the research designs of these evaluation studies do not preclude the alternative interpretation that motivation and achievement were more easily modified by early intervention than cognitive abilities. Interestingly, however, the data do demonstrate that certain child characteristics (e.g., age, initial IQ level) moderate the effects of program emphases.

The findings with respect to the influence of teacher and program variables on developmental change shed less light on the issue of specificity versus equality of effects. Whether specific program approaches are associated with the corresponding types of outcomes or with a variety of outcomes is not clearly resolved by these findings. Global appraisals of the first year's findings of the Planned Variation study (see Bissell, 1971) reported comparable gains across all well-implemented curricula models. The largest differences in effects were found not between models, but between well-implemented curricula

models and more loosely formulated traditional Head Start programs. Bissell (1971) noted that fine grained analyses of the well-implemented programs did show specific but small effects corresponding to particular program emphases. Since most of the Head Start programs included in the 1966-1969 evaluation studies were loosely formulated, traditional Head Start programs, the finding of specific effects is misleading; the primary question of whether a given program emphasis was more effective or merely better implemented cannot be answered by these data.

Specific program/child combinations. The last factor to be considered as a significant influence on performance gains are the program/child interactions -- specific groups of children who benefitted from particular types of program approaches. The interactions reported in these studies are few in number and small in size compared with the effects of program variables when considered independent of child variables. For this reason the interactions should be construed as suggestive rather than prescriptive of different educational treatments for different kinds of children. 18/

In her description of two hypothetical types of early intervention programs, Stearns (1971) distinguishes between traditional preschool programs and experimental research projects. She concludes that the former produce smaller gains and more interactions between individual and program characteristics, a tendency born out in these evaluation findings and consistent with the formulation of Head Start as a broadly based child development program.

Implications for Early Childhood Intervention

In this section of the report the implications of these evaluation findings for early childhood intervention research and programs will be presented and discussed. The order of presentation will parallel the preceding discussion of the findings.

- Research and program model development should be directed at exploring alternative strategies of family-centered

18/ The interpretation of the interactions in this study deserves additional comment. It must be noted that the interactions were uncovered by an analytic model which was especially sensitive to their occurrence and which may have generated many "false positives." Nonetheless, the interactions were consistent with the findings for individual characteristics taken one at a time. It is unfortunate that more replication of these important findings was not possible. The use of somewhat different measures across the three years of study and of somewhat different statistical techniques across the two reports, makes the number of direct replications relatively small.

intervention. This recommendation is supported by the findings that the family is a major determinant of what the child brings to the program, which in turn influences the benefits he or she derives from the program. These particular findings also point out characteristics of the parent-child relationship which are related to differences in child competence upon entry and which are potentially manipulable through family-intervention strategies, e.g., parental discipline techniques, frequency of reading to the child, adult accessibility. These findings also suggest that child-oriented interventions should explore new ways to build upon and enhance parents' skills as effective educational and socializing agents.

- Children from a wide variety of circumstances are likely to benefit in the short term from early intervention programs like Head Start. The findings reported here show that virtually all children gain from their Head Start experience, although some children gain more than others or gain on different outcome variables. It seems likely that the broadly-based and multi-aimed objectives of most Head Start programs provide a sufficient diversity of experience to match some developmental needs of most of the children attending the program. The findings illuminate the importance of systematic assessment of the developmental needs of the entering population of Head Start children with a view toward improving the individualization of Head Start experiences.

- There is no one best program or curriculum approach for all children; program approaches should be matched to the child and the teacher. Within the range of program variations and children studied in these evaluations, no single program approach proved to be superior for all children across all outcome domains. What a child brings to a program influences the way he is affected by his Head Start experiences. Further, the experiences which influence his cognitive development may be different from the experiences which influence his personal-social development. This means that studies averaging gains over classes and children, or focusing on only one kind of outcome are likely to misrepresent the most significant effects. In terms of early educational practices, the findings imply a need to develop techniques for fine-grained developmental assessment, to verify the effectiveness of individualized programs, and then to develop feasible ways of providing such programs on a large scale, if a focus on early child development continues to be a national priority.

In terms of early intervention research, the findings imply a need for more investigation of the interactions between teacher characteristics (including teacher style) and child characteristics, and more precise identification of teacher

skills and attributes which are likely to enhance children's gains. It should be noted that the Child Development Associate effort is currently addressing some of these issues.

- Improved program planning, more precise articulation of program objectives, and closer approximation of practice to child development theory, should improve the matching of program experiences to the needs of individual children and should increase overall program effectiveness. The interactions between child and program characteristics and the specificity of effects for certain program elements underscore the importance of more rigorous program planning on the one hand, and the assessment of children's needs, on the other. The findings with respect to child/program interactions also point to the continuing need to achieve congruence between program objectives, child development outcomes, and objective outcome measures.

Summary

As a concluding reminder, the reader is urged to consider the findings from these national evaluations in the context of the objectives of Project Head Start, as stated in the Cooke Memorandum. These objectives are heavily weighted in favor of social and emotional growth; however, the most significant relationships reported here are weighted in favor of achievement motivation, skills, and cognitive growth. Furthermore, of the many components of Head Start, only one (center classroom) was used as a data basis for these evaluations.

Apart from the obvious need to develop measurement instruments appropriate for non-cognitive objectives, this "mismatch" between program objectives and assessment may illuminate the possibility that a broad program can never be fairly evaluated in its entirety. Just as these data have shown that some centers do better at advancing cognitive development, it is likely that some centers have more successful career development and training programs than others. No single evaluation can ever capture all that is Head Start.

REFERENCES

Adkins, D.C. and Ballif, B.L. Motivation to Achieve in School. Final report, Head Start Evaluation and Research Center, University of Hawaii, January, 1970.

Anderson, R.H. and Shane, H.G. (Eds.) As the Twig is Bent: Readings in Early Childhood Education. Boston: Houghton-Mifflin Company 1971.

Auleta, M. (Ed.) Foundations of Early Childhood Education: Readings. New York: Random House, 1969.

Averich, S.C. and Donaldson, T. What Do We Know About Educational Effectiveness? Santa Monica, California: Rand Corporation, 1970.

Bane, M.J. and Jencks, C. The Schools and Equal Opportunity. Saturday Review of Education, September 16, 1972.

Bates, B. Head Start 1965-1967: A Descriptive Report of Programs and Participants. Washington, D.C.: Office of Child Development, Department of Health, Education and Welfare, 1969.

_____; Head Start 1968: A Descriptive Report of Programs and Participants. Washington, D.C: Office of Child Development, Department of Health, Education and Welfare, 1970.

_____; Head Start 1969-1970: A Descriptive Report of Programs and Participants. Washington, D.C.: Office of Child Development, Department of Health, Education and Welfare, 1972.

Beller, E.K. The evaluation of effects of early educational intervention on intellectual and social development of lower-class disadvantaged children. In E. Grotberg (Ed.), Critical Issues in Research Relating to Disadvantaged Children. Princeton, New Jersey: Educational Testing Service, 1969.

Bereiter, C. and Engelmann, S. Academic Preschool, Champaign, Illinois. "It Works" Series, Preschool Programs in Compensatory Education. Washington, D.C.: Office of Education, Department of Health, Education and Welfare, 1969.

Bissell, J.S. The Cognitive Effects of Preschool Programs for Disadvantaged Children. Washington, D.C.: National Institute of Child Health and Human Development, Department of Health, Education and Welfare, 1970.

_____: Implementation of Planned Variation in Head Start. I. Review and Summary of the Stanford Research Institute Interim Report: First Year of Evaluation. Washington, D.C.: Office of Child Development, Department of Health, Education and Welfare, 1971.

Bloom, B. Stability and Changes in Human Characteristics. New York: John Wiley and Sons, 1964.

Boyd, J. Project Head Start, Summer 1966. Facilities and Resources of Head Start Centers. Princeton, New Jersey: Educational Testing Service, 1966.

Braun, S.J. and Edwards, E.P. History and Theory of Early Childhood Education. Worthington, Ohio: Charles A. Jones, 1972.

Bronfenbrenner, U. Volume II, Is Early Intervention Effective? A Report on Longitudinal Evaluations of Preschool Programs. Washington, D.C.: Office of Child Development, Department of Health, Education and Welfare, 1974.

Butler, A.L. Current Research in Early Childhood Education: A Compilation and Analysis of Program Planners. Washington, D.C.: American Association of Elementary-Kindergarten-Nursery Education, 1970.

Campbell, D.T. and Erlenbacher, A. How Regression artifacts in quasi-experimental evaluation can mistakenly make compensatory education look harmful. In Hellmuth, J. (Ed.), Disadvantaged Child, Volume 3, New York: Brunner/Mazel, Inc., 1970.

Cicarelli, V.G., Cooper, W.H. and Granger, R.L. The Impact of Head Start: An Evaluation of the Effect of Head Start on Cognitive and Affective Development: Volume I, Text and Appendices A to E. (Report on Contract B89-4536, Westinghouse Learning Corporation and Office of Economic Opportunity). Washington, D.C.: Office of Economic Opportunity, June, 1969.

Coleman, J. Equality of Educational Opportunity. Washington, D.C.: U.S. Government Printing Office, 1966.

Cooke, R. Recommendations for a Head Start Program by A Panel of Experts. Washington, D.C.: Office of Economic Opportunity, 1965.

Costello, J. and Binstock, E. Review and Summary of A National Survey of the Parent Child Center Program. Washington, D.C.: Office of Child Development, Department of Health, Education and Welfare, 1970.

Datta, L. A Report on Evaluation Studies of Project Head Start. Washington, D.C.: Office of Child Development, Department of Health, Education and Welfare, 1969.

DiLorenzo, L.T., Salters, R. and Brody, J. Prekindergarten Programs for Educationally Disadvantaged Children. (Final Report, Project No. 3040, Contract No. OE6-10-040). New York: New York Education Department, 1969.

Dittman, L., et.al. Study of Selected Children in Head Start Planned Variation, 1969-1970. (Final Report, Office of Child Development). Institute for Child Study, University of Maryland, March, 1971.

Educational Testing Service. Disadvantaged Children and Their First School Experiences: Preliminary Description of the Initial Sample Prior to School Enrollment. Princeton, New Jersey: Educational Testing Service, July, 1971.

Emmerich, W. Structure and Development of Personal-Social Behaviors in Preschool Settings. (Report on Grant H-8256, Office of Child Development and Educational Testing Service). Princeton, New Jersey: Educational Testing Service, November, 1971.

Evans, E.D. Contemporary Influences in Early Childhood Education. New York: Holt, Rinehart and Winston, Inc., 1971.

Gray, S.W. and Klaus, R.A. The Early Training Project: A Seventh Year Report. Journal of Child Development, 1970, 41, 908-924.

Grotberg, E.H. Review of Research, 1965 to 1969. (Project Head Start, Office of Economic Opportunity Pamphlet 6108-13, ERIC no. ED028308). Washington, D.C.: U.S. Government Printing Office, 1969.

Project Head Start. Head Start Guidelines. Washington, D.C.: Office of Economic Opportunity, 1967.

Hellmuth, J. (Ed.) Disadvantaged Child. Seattle, Washington: Special Child Publications, 1967.

_____: (Ed.) Disadvantaged Child, Volume 3, Compensatory Education: A National Debate. New York: Brunner/Mazel, Inc., 1970.

Hodges, W., McCandless, B.R. and Spicker, H. The Development and Evaluation of A Diagnostically Based Curriculum for Preschool Psycosocially Deprived Children. (Final Report to the U.S. Office of Education). Washington, D.C.: Office of Education, Department of Health, Education and Welfare, 1967.

Hoefner, R., Stern, C., and Nummedal, G. CSE-ECRC Preschool/Kindergarten Test Evaluations. Los Angeles, California: UCLA Graduate School of Education, 1971.

Hotelling, H. The selection of variates for use in prediction with some comments on the general problem of nuisance parameters. Annual on Methods of Statistics, 1940, 11, 271-283.

Hunt, J. McV. Intelligence and Experience. New York: The Ronal Press Company, 1961.

_____: (Chairman) A Bill of Rights for Children. Report of the White House Task Force on Early Childhood, Washington, D.C., 1966.

Jensen, A. How Much Can We Boost IQ and Scholastic Achievement? Harvard Educational Review, 1969, 39, 1-123.

Karnes, M. A Research Program to Determine the Effects of Various Preschool Intervention Programs on the Development of Disadvantaged Children and the Strategic Age for Such Interventions. (ERIC no. ED017319). Urbana, Illinois: University of Illinois, 1968.

_____: The Ameliorative Preschool Program Champaign, Illinois. "It Works" series, Preschool Programs in Compensatory Education. Washington, D.C.: Office of Education, Department of Health, Education and Welfare, 1969.

Kirschner Associates, Inc. A National Survey of the Impacts of Head Start Centers on Community Institutions. (Contract No. B89-4638, Project Head Start), Washington, D.C.: Office of Child Development, Department of Health, Education and Welfare, May, 1970.

Klein, J. and Datta, L. Head Start Planned Variation Study. Washington, D.C.: Office of Child Development, Department of Health, Education and Welfare, November, 1971.

Kohlberg, L. and Mayer, R.S. Preschool Research and Preschool Objectives: A Critique and a Proposal. Unpublished manuscript, Harvard University, 1971.

Light, R.J. and Smith, P.V. Accumulating evidence: Procedures for resolving contradictions among different research studies. Harvard Educational Review, 1971, 41, No. 4, 429-471.

Miller, L., et.al. Four Preschool Programs: Their Dimensions and Effects. (Progress Report No. 10, Office of Child Development). Washington, D.C.: Office of Child Development, Department of Health, Education and Welfare, May, 1971.

Mostellar, F. and Moynihan, P. (Eds.) On Equality of Educational Opportunity. New York: Random House, 1972.

Nimnicht, G.P. Overview of a Responsive Program for Young Children. Unpublished Manuscript from Far West Regional Laboratory for Educational Research and Development, Claremont Hotel, Berkeley, California, 1970.

Planning Research Corporation. Results of the Summer 1965 Project Head Start. (Report of Contract OEO-753). Washington, D.C.: Office of Economic Opportunity, May, 1966.

_____: A Study of the Full Year 1966 Head Start Programs. Washington, D.C.: Office of Economic Opportunity, 1966.

Research Triangle Institute (G. Dunteman, Project Director). A Report on Two National Samples of Head Start Classes: Some Aspects of Child Development of Participants in Full Year 1967-1968 and 1968-1969 Programs. (Final Report on Contract HEW-OS-70-207). Washington, D.C.: Office of Child Development, Department of Health, Education and Welfare, July, 1972.

Rohwer, W.D. Prime time for education: Early childhood or adolescence? Harvard Educational Review, 1971, 41, 316-341.

Rowe, R.R., Fein, R.A., Butler, J.A., Rowe, M.P., Bunting, C.I. and Johnson, G.R. Child Care in Massachusetts: The Public Responsibility. A Study for the Massachusetts Advisory Council on Education, Harvard University, February, 1972.

Ryan, S. (Ed.) Volume I, Longitudinal Evaluations. A Report on Longitudinal Evaluations of Preschool Programs. Washington, D.C.: Office of Child Development, Department of Health, Education and Welfare, 1974.

Shipman, V. Disadvantaged Children and Their First School Experiences. ETS-Head Start Longitudinal Study. Princeton, New Jersey: Educational Testing Service, February, 1970.

Skeels, H.M. Adult status of children with contrasting early life experiences: A follow-up study. Monographs of the Society for Research in Child Development, 1966, 31, No. 3 (Serial No. 105).

Skeels, H.M. and Dye, H.B. A study of the effects of differential stimulation on mentally retarded children. Proceedings and Addresses of the American Association on Mental Deficiency, 1939, 44, 114-136.

Smith, M.S. and Bissel, J.S. Report analysis: The impact of Head Start. Harvard Educational Review, 1970, 40, 51-104.

Sprigle, H.A. Learning to Learn Program, Jacksonville, Florida. "It Works" series, Preschool Programs in Compensatory Education. Washington, D.C.: Office of Education, Department of Health, Education and Welfare, 1969.

Stanford Research Institute. Interim Report: First Year of Evaluation. Implementation of Planned Variation in Head Start: Preliminary Evaluations of Planned Variation in Head Start According to Follow-through Approaches. Menlo Park, California: Stanford Research Institute, May, 1971.

Stearns, M.S. Report on Preschool Programs: The Effects of Preschool Programs on Disadvantaged Children and Their Families. Washington, D.C.: Office of Child Development, Department of Health, Education and Welfare, 1971.

Stodolsky, S.S. Defining treatment and outcome in early childhood education. In H. Walberg (Ed.), Rethinking Urban Education. San Francisco, California: Josey-Bass, Inc., 1972.

System Development Corporation (J.E. Coulson, Project Director). Effects of Different Head Start Program Approaches on Children of Different Characteristics: Report on Analysis of Data from the 1968-1969 National Evaluations. (Final Report on Contract HEW-OS-70-168). Washington, D.C.: Office of Child Development, Department of Health, Education and Welfare, May, 1972(a).

_____: Effects of Different Head Start Program Approaches on Children of Different Characteristics: Report on Analysis of Data from the 1966-1967 and 1967-1968 National Evaluations. (Technical Memorandum on Contract HEW-OS-70-168). Washington, D.C.: Office of Child Development, Department of Health, Education and Welfare, June, 1972(b).

Van de Riet, V. A Sequential Approach to Early Childhood and Elementary Education, Phase III. (Report on Grant OCD-H-8222 B/H/O, Office of Child Development). Washington, D.C.: Office of Child Development, Department of Health, Education and Welfare, January, 1972.

Walker, D.K. Socioemotional Measures for Preschool and Kindergarten Children. San Francisco, California: Josey-Bass, Inc., 1973.

Wargo, M.J., Campeau, P.L. and Tallmadge, G.K. Further Examination of Exemplary Programs for Educating Disadvantaged Children. Palo Alto, California: American Institutes for Research in the Behavioral Sciences, July, 1971.

Weber, G. Inner City Children Can Be Taught to Read: Four Successful Schools. Occasional Papers, Number Eighteen. Washington, D.C.: Council for Basic Education, 1971.

Weikart, D.P. A comparative study of three preschool curricula. Paper presented at the Society for Research in Child Development, biennial meeting, Santa Monica, California, March, 1969.

_____: The Ypsilanti Preschool Curriculum Development, 1968-1971. Ypsilanti, Michigan: High-Scope Educational Research Foundation, 1971.

Zigler, E. and Butterfield, E.C. Motivational aspects of changes in IQ test performance of culturally deprived nursery school children. Child Development, 1968, 39, 1-14.

APPENDIX A

Measure Selection and Psychometric Characteristics

Child and Family Measures 1968-1969

Information available on the child included pre- and post-test scores on the following instruments: Stanford-Binet, Inventory of Factors Affecting the Stanford-Binet, Hertzig-Birch Response Style, Gumpgookies (achievement motivation). Sociometric Play Situation, Animal House test form the Wechsler Preschool and Primary Scale of Intelligence, and the Caldwell-Soule Preschool Inventory. Extensive demographic information was available for all children on a master data card. This included the child's age, sex, geographic region, urbanicity, ethnicity, and so forth.

1. Cognitive Development Measures

Stanford-Binet IQ. The Stanford-Binet was developed as an intelligence test, as a measure of the individual's ability to learn from experience. This interpretation requires that the individual tested have been exposed to experiences similar to those of individuals on whom the measure was standardized. In the 1957 revision of this test, the standardization sample over-represented white and more advantaged children. Even if the sample had been representative of the ethnic and economic circumstances of children in the general population, the performance of any group whose life experiences diverged from the norm sample as a whole, would have to be interpreted as an indication of both experience per se and the ability to learn from experience. The higher IQs of more advantaged children do indicate a greater ability to perform certain tasks, but that ability is related both to possible differences in learning capacity and to certain differences in life circumstances that favor performance on such test items. The lower IQs typically achieved by children from economically constricted circumstances indicate less ability on the particular skills measured by the Binet, but again, the level of performance is due both to possible differences in learning capacity and to certain differences in life experiences that do not favor performance on test items. In addition, scores on this and other individually administered tests depend heavily on the child's cooperativeness, motivation, and rapport with the examiner.

As a measure of performance, the absolute level of the

Stanford-Binet scores has repeatedly been found sensitive to intervention of many kinds; the relative position of individual scores within a group exposed to the same intervention changes much less. Where the intervention is very effective, the individual differences within groups may actually be increased; as children become more able to learn they in turn respond to new opportunities for learning. The Binet frequently has been reported to predict school achievement as well as rate of learning on a variety of new tasks.

Animal House. This subscale of the Wechsler Preschool and Primary Scale of Intelligence is similar to the adult Digit-symbol or Coding tests. The materials consist of a board with pictures of four different animals arranged in rows with spaces for "houses" beneath each picture, and a supply of colored pegs -- the "houses". A key at the top of the board identifies the color for each animal's house, and the child's task is to correctly place a house under each animal on the board. The raw score is a combination of time, errors, and omissions; an age-scaled score may also be derived. The WPPSI is a standardized test of the ability to learn, and the Animal House subtest was selected as a relatively culture-free, nonverbal test of the ability to learn a new task.

Preschool Inventory. The Caldwell-Soule Preschool Inventory was developed in 1965 specifically to assess the level of cognitive maturation of Head Start children. It seems to measure cognitive functioning at a more concrete level than the Stanford-Binet, and is interpreted as a measure of achievement or school readiness. There are four subtests to the scale: Social Responsiveness, Associative Vocabulary, Concept Activation/Numerical, and Concept Activation/Sensory. Scores on the subscales are highly intercorrelated. Although the form used in 1968-69 was then an experimental version, it is now a standard, commercially available instrument. Data from the 1968-69 pretesting provided the national standardization data for low-income children.

2. Social-Emotional Development Measures

Inventory of Factors Affecting the Stanford-Binet. This instrument is a modification of the fact sheet of the Stanford-Binet record form. The examiner rates the child on his behavior during the Stanford-Binet testing session, assessing such traits as compulsivity, concentration, and activity level. This adaptation was prepared by Dr. Herbert Zimiles of the Bank Street College of Education.

Hertzig-Birch Response Style. This instrument, like the Inventory of Factors Affecting the Stanford-Binet, is given in conjunction with the Stanford-Binet, and was designed to measure various cognitive styles that the child employs when responding to test items. For example, a distinction is made between a wrong response where the child "works at" a problem solution and a wrong response where the child does not. Hertzig, et.al. (1968) have demonstrated differences in these cognitive styles among various ethnic and socio-economic groups.

Gumpgookies. The Gumpgookies test is an experimental measure of achievement motivation designed for preschool children (Adkins and Ballif, 1970). The form used for initial testing had 100 items; the post-test version had only 55 items. The reduction in length was required both by the short attention span of the children and by psychometric shortcomings of some items. Each item consists of a picture of two pillow-like creatures, one of which is doing something that is more achievement oriented than what the other one is doing. For example, one figure might be successfully completing the building of a block tower while the other was sitting in a heap of fallen blocks. The child is asked which Gumpgookie is his Gumpgookie (i.e. the one that acts like he does). The internal consistency reliability for the test is .76. Although Gumpgookies scores do discriminate between children rated by their teachers as high or low in motivation, the scores are also influenced by idiosyncrasies in responding, such as position preference.

Sociometric Play Situation. In this task designed by Dr. Robert Boget, each child selects (a) three play situations depicted on six cards, and then (b) the photograph of a preferred classmate to put into each of the three play situations. Thus, measures of popularity and isolation can be derived for each child. The structure of the task resulted in artifactual pre-post correlations: if some children's relative popularity increased, the remaining children's relative popularity had to decrease.

3. Parent and Family Measures

Parent Interview. Pre- and post-interviews with the parent (usually the mother) yielded information about potential variables including: the family's socioeconomic status (family income, a father's occupation, mother's education, father's education); aspects of family life (family structure, frequency of reading to the child,

amount of living space, parent participation in the Head Start program); parents' attitudes toward society in general (expectations for the child's education, aspirations for the child's education, feelings of alienation). In addition, the interview presented many open-ended questions intended to reflect typical styles of social interaction between parent and child. The variables and measures pertaining to parents' attitudes and parent-child interaction merit greater elaboration below.

The measures of parental attitudes toward society and life in general are interpreted as "locus-of-control" measures. The individual with a sense of internal control believes that he has the power to control his own life and to influence the lives of others; the individual with a sense of external control believes he has little command over what happened to him and his family. Scores on tests of locus-of-control have been found to predict which individuals from disadvantaged circumstances will be able to overcome their environmental constrictions. For example, Coleman (1965) found a positive correlation between high test performance and internal control in Black children; in other studies, changes in locus-of-control have been associated with changes in achievement. Nonetheless, the direction of a causality is an unresolved question: are "internals" more accurate in rating their influence and power because they really are more competent, or are they able to take advantage of opportunities simply because they believe they can?

Information about parent-child interactions was derived from four questions about the parents' disciplinary practices. One question asked the mother to name one of the worst things that her Head Start child does and explain how she responds to it. The second question asked about "little annoying things" and the mother's response; questions three and four related to child behaviors which the mother liked and liked "a great deal." The responses were categorized according to the means by which parents modified or reinforced the child's ongoing behavior. The broad categories were: non-intervention, intervention, qualified power assertion, and unqualified power assertion. This measure was developed by Dr. Irving Siegle and Dr. Bela Fehrer, with Dr. Fehrer supervising the coding of the present data. Similar items have been found to predict preschool achievement.

Another open-ended question asked the mother was, "What would you tell your child to do on the first day of school?" This item measures how the mother views the

teacher, classroom, social situations, achievement, home preparation, and personal safety. The mother's responses were categorized by trained raters under the direction of Dr. Robert Hess. These ratings have been shown to be sensitive to social class differences within samples of Black children, and to predict both preschool and later primary school achievement.

Teacher and Program Measures 1968-1969

1. Teacher Self-Reports

Teacher Interview. Questions on this instrument explored program focus, importance of child development goals, description of teaching approach, and kinds of commercial curriculum materials used by the teacher. It was developed as an independent reliability measure for this intervention studies of 1968-69 by Dr. Carolyn Stern.

Checklist of Administrative Variables. This checklist was designed, to determine the person(s) responsible for each of the 31 key functions or activities in full-year Head Start programs. Examples of these functions included preparing the agenda for teachers' meetings and planning parent orientation sessions. The checklist was prepared by Dr. Shuell Jones as a measure of the quality of program administration as seen by Head Start teachers.

Characteristics of Teaching Staff Form. This questionnaire, prepared by Dr. Carolyn Stern, was administered at the end of the program year and provided information regarding the head teacher, teacher, and teacher aide: age, sex, ethnicity, education, training, experience.

Class Facilities and Resources Inventory. This form is similar to that used in census surveys to determine compliance with Head Start guidelines. The class information included the operating length of the program, length of class day, number of children in class, and the various educational and physical resources available to the children.

2. Observations of Classroom Activities

Post Observation Teacher Rating Scale. This instrument consisted of 33 ratings of teacher behavior made by the

same observer and on the same day that OSCI observations were obtained (see below). Ratings included such teacher behavior as her reliance on on-going activities, her attention to groups, and her awareness of pupil frustration. This scale, developed by Dr. John Dopyera, is based on the work of Harvey, Hunt, and Schroeder (1967) on the "openess" and "closedness" of individual style.

Observation of Substantive Curricular Input (OSCI). The OSCI was developed by Dr. Carolyn Stern. The instrument consists of a complex schedule that is used by a trained observer to code the behavior of the class and the teacher.

The OSCI coding system is based on a series of three minute scans of on-going activity in the classroom. In some of the scans the observer looks at what the children are doing, and in others she looks at the teacher and teacher-aide. For each scan, the behaviors of the largest group of children are observed and coded. The major codes reflect the context of activity (i.e., building), the content of the activity (i.e., mechanical), the locus-of-behavior (i.e., teacher or child initiated), and teacher involvement (i.e., present and watching). Within the three-minute period, the observer then locates the next largest group, and repeats the coding process. A three-minute record could consist of one unit (if all the children are occupied in one activity) or of as many units as there are children (if each child is doing something individually). Three-minute scans of teacher activity record teacher context, teacher content, and teacher involvement for six 30-second intervals.

These teacher and classroom observations were collected on five separate days spread through the middle and end of the program year. Each daily observation covered most of the school day for the majority of Head Start sample classes. Observers attended special three-day training sessions and were supervised regularly for quality and reliability.

These voluminous data were analyzed by concentrating on selected content-context categories and control categories for both class and teacher observations. Frequency measures were then intercorrelated, factor analyzed, and factor rotated. Six rotated factors were retained for further analyses in both the class and teacher data.

The six teacher factors were identified as follows:

I. Social-emotional Interaction. Partially defined by emotional and social behavior of the teacher.

II. Structured Lessons -- Large Group. Partially defined by use of structured language materials, structured visual discrimination lessons, use of program materials.

III. Art Activities. Partially defined by art content in a painting context, and verbal communication in an art context.

IV. Creative Instruction -- Small Group. Partially defined by visual motor activities, small groups, presence of large and small muscle equipment.

V. Routines. Partially defined by teacher use of mechanical devices, and teacher activity in routine context.

VI. Receptive Learning. Partially defined by language content in a watching-listening context, visual discrimination in an activity context, and presence of language materials.

The six OSCI class factors were labelled as follows:

I. Structured Lessons. High loadings from presence of programmed materials, language through structured lessons, control by the head teacher, and large group activities.

II. Group Activities and Routines. Partially defined by music, drama, and art activities in a social context.

III. Social-emotional Interaction. Partially defined by emotional content in all contexts, social content in a physical contact context, and control by a child.

IV. Verbal Communication. Primarily defined by verbal communication in a variety of contexts including activities, arts, and routines.

V. Instruction in Creative Arts. Primarily defined by visual motor content in an activities context, child control, small groups, presence of dramatic play equipment, and dramatic content in a performing context.

VI. Language and Discrimination Learning. Primarily defined by the presence of language materials, language content in a watching-listening context, and visual discrimination in an activities context.

Measures Used in 1967-1968 and 1966-1967

The data collected in 1967-1968 and 1966-1967 are used in this summary report chiefly for the purpose of indicating replication of findings reported for 1968-1969. Therefore, the discussion of the measuring instruments for 1967-1968 and 1966-1967 is briefer and more general than for those used in the primary data base, i.e., 1968-1969.

1. Child and Family Measures

Stanford-Binet IQ. These scores were available for most children in the 1966-1967 and 1967-1968 samples.

Inventory of Factors Affecting the Stanford-Binet. Scores were available for 1967-1968 only.

Preschool Inventory. Scores were available for 1966-1967 only.

Head Start Behavior Inventory. Behavior ratings made by the Head Start teachers were available for 1966-1967 only.

Parent Interview. Pre- and post-program interviews similar to those obtained in 1968-1969 were available for both 1966-1967 and 1967-1968.

2. Teacher and Program Measures

Characteristics of Teaching Staff Form. This instrument provided data similar to that of the 1968-1969 study for the two previous years.

Center and Classroom Composition Instrument. This questionnaire provided information similar to that from the Class Facilities and Resources Inventory.

Observation of Substantive Curricular Input (OSCI). The 1967-1968 OSCI was similar to the 1968-1969 OSCI but gathered information only on the class, rather than both class and teacher. Analysis of the intercorrelations of the frequencies of the various class content-context categories across classes yielded only four rotated Class Factor Scores, rather than six, as in 1968-1969. These

four factors are briefly described below:

I. High Cognitive Activity in a low structure situation. Mostly defined by high loadings of verbal instruction during learning activities and verbal instruction during routines.

II. Routines and Rules. Characterized by high loadings on rules emphasized during routines, whole group activities, and social interaction during learning activities.

III. High Cognitive Activity in a High Structure Situation. Partially defined by positive loadings of visual discrimination and cognitive input during watching or listening, and a negative loading for individual activity.

IV. Child-Centered, Unstructured. Partially defined by high positive loadings for social interaction during learning activities and high negative loadings for mechanical performance of routines.

Social Interaction Observation (SIO). This instrument, unique to the 1967-1968 data, was used to gather detailed information about child-child and adult-child interactions in the classroom. Developed at the University of Kansas, the SIO records the number of verbal and nonverbal initiations from the child to an adult, the number of responses by the adult, and the number of adult initiations responded to by the child.

Observer Rating Form. This instrument was used only in 1966-1967 to obtain classroom information. It was developed by Dr. John Pierce-Jones and his associates.

APPENDIX B

Selection of Variables for Analysis

RTI

Originally, 90 explanatory and outcome variables were considered for analysis based on previous research and experience. Variables from the preliminary list were subsequently dropped for a variety of reasons, the most common ones being: (1) computational and cost factors in deriving an index from the raw data provided on tape, (2) undesirable marginal distributions (e.g., no variation in an explanatory variable), (3) low internal consistency reliability of derived or scaled variables as measured by Kuder-Richardson Formula 20, and (4) apparent uninterpretability or extreme redundancy as shown in analyses of the intercorrelations and interdependencies among many sets of preliminary outcome and explanatory variables.

SDC

The method selected for the screening process was a correlational analysis approach which permitted examination of large numbers of variables at relatively low cost. For each dependent variable, the Prearson product-moment correlation was computed between the pre-test score and each program variable; next the correlation was computed between the post-test score and each program variable. The differences between those two correlations were computed and a "t" test was performed on each pair to test the null hypothesis of "no difference." A correlation increase was interpreted as suggestive of a positive relationship between the program variable and the performance measure, and a decrease as a negative relationship. After many variables were eliminated from consideration because the pre-post correlation difference did not reach statistical significance, a few were returned to the list because of their importance in the literature on their place in a priori hypotheses.

Selection of Analytic Models

RTI

To identify effects that were relatively independent of subsetting variables, RTI performed a series of four-way ANOVAs in which the children's adjusted gain scores were examined when all possible oneway effects and two-way interactions (except those of the prime oneway variable to be tested) were taken into account. This approach is maximally sensitive to the stability of the main effect when several subsetting variables are simultaneously taken into account. RTI

also reported results from multiple regression analyses of classroom and teacher characteristics (as measured by the OSCI) used to predict children's gains on the Stanford-Binet, Preschool Inventory, and Feelings of Inadequacy Scale.

SDC

Interactions were inferred from the retention or drop-out of the statistical significance of a main effect when tested in separate subset analyses (for example, high vs. low levels of "Large Muscle Equipment" was tested separately for groups defined by initial IQ score low, medium, and high). In all of these ANOVAs (over 1000), a specially adjusted post-test score was used. In effect, the initial score was a covariate, except that the regression weight of the lowest level of the program variable was used for all of the calculations. This procedure was chosen as appropriate for an exploratory study, in spite of the fact that it does increase the chance of finding spuriously significant interactions.

Differences in Application of Analytic Models

1968-1969 Data

1. Child Characteristics

Intelligence (Stanford-Binet IQ)

RTI. Used as an outcome variable.
SDC. Used as an outcome variable, and used pretest Stanford-Binet score as a subsetting variable for testing interactions.

Ability to Learn a New Task (Animal House)

RTI. Used raw score as an outcome variable.
SDC. Used scaled scores as an outcome variable.

Achievement and School Readiness (Caldwell-Soule Preschool Inventory)

RTI. Used raw score on total PSI as an outcome variable.
SDC. Used raw scores on the four subscales and an age-normed score on the total score as outcome variables.

Adjustment to a Novel Situation (Inventory of Factors Affecting the Stanford-Binet

RTI. Used three rationally derived subscales as outcome variables: Behavior Problems, Motivational Problems, and Feelings of Inadequacy.

SDC. Used total score as an outcome variable interpreted as an overall index of adjustment.

Cognitive Style (Hertzig-Birch Response Style)

RTI. Did not use these data.
SDC. Used three subscores as outcome variables: Work Response Score, Verbal Response Score, and Spontaneous Response Score.

Achievement Motivation (Gumpgookies)

RTI. Used as an outcome variable
SDC. Did not use these data.

Children's Sociability (Sociometric Play Situation)

RTI. Did not use these data.
SDC. Used a measure of social isolation as an outcome variable.

2. Parent and Family Characteristics

Socio-economic Status, Parental Attitudes, Styles of Parent-Child Interactions, Household Routines (Parent Interview)

RTI. Used these as predictor variables.
SDC. Used some family characteristics as outcome variables and others as predictor variables.

3. Teacher and Program Characteristics

Program Emphasis, Goals, and Teaching Approach (Teacher Interview)

RTI. Derived several program variables from the teacher responses.
SDC. Used these responses as a primary indicator of program content.

Administrative Variables (Checklist)

RTI. Made no use of these data.
SDC. Made no use of these data.

Teacher's Background (Characteristics of Teaching Staff Form)

RTI. Used these variables as predictors.
SDC. Used these variables as predictors.

Program Facilities and Resources (Class Facilities and Resources Inventory)

RTI. Used four (rather global) categories as predictor variables.

SDC. Used two (rather specific) categories as predictor variables.

Teacher's Classroom Behavior (Post Observation Rating Scale)

RTI. Used some items as predictor variables.
SDC. Used some items as predictor variables.

Teacher's and Children's Classroom Behaviors (Observation of Substantive Curricular Input)

RTI. Used these data as predictor variables and as subsetting variables for testing interactions.

SDC. Did not use these data.