DOCUMENT RESUME

ED 205 279 PS 012 290

AUTHOR	Berrueta-Clement, John: And Others
TITLE	Assessment of Program Impact Through First Grade, Volume V: Impact on Children. An Evaluation of Project Developmental Continuity. Interim Report V.
INSTITUTION	High/Scope Educational Research Foundation, Ypsilanti, Mich.
SPONS AGENCY	Office of Human Development Services (DHEW), Washington, D.C.
PUB DATE	Dec 80
CONTRACT	105-78-1307
NOTE	201p.: For other volumes in this report, see PS 012 286-291.
EDRS PRICE	MF01/PC09 Plus Postage.
DESCRIPTORS	*Academic Achievement: *Child Development: Control Groups: *Cooperative Programs: Coordination: Disabilities: *Disadvantaged Youth: Educational Change: Emotional Development: Family Characteristics: Grade 1: *Health: Language Acquisition: Longitudinal Studies: Nutrition: Preschool Education: Primary Education: *Program Effectiveness: Program Evaluation: Social Development: Spanish Speaking: Teacher Characteristics
IDENTIFIERS	*Project Developmental Continuity: Project Head Start

ABSTRACT

Fifth in a series of six volumes reporting outcomes of the preliminary evaluation of an educational intervention, this report presents the findings of the effects of Project Developmental Continuity (PDC) up to the time the evaluation study's cohort of children completed grade 1. Preliminary findings concerning the relationship between variables in the teacher, parent and child domains are also provided. Begun at 15 sites in 1974 with the purpose of ensuring disadvantaged children receive continuous individualized attention as they progress from Head Start through the early primary grades, PDC emphasizes the involvement of school personnel and parents in formulating educational goals and in curriculum development. Chapter I of this volume presents a brief history of the PDC program. Chapter II describes the conceptual framework guiding the evaluation of PDC processes and effects on children. Data collection and analysis procedures are discussed in Chapter III. Chapter IV presents the descriptive findings regarding the sample and the characteristics of the instruments. Chapter V describes the results of the analysis of the child outcome measures. Chapter VI summarizes the major findings and discusses implications for the future. Appendices A and B provide, respectively, descriptions of the child development measures used and a comparison of summary score distributions by treatment group. (Author/RH)

Project Developmental Continuity Evaluation

Interim Report X: Assessment of Program Impact Through First Grade

Vol. V
Impact on Children

December 1980, High/Scope Educational Research Foundation

This report was prepared for the Early Childhood Research and Evaluation Branch, Administration for Children, Youth and Families, Office of Human Development Services, Department of Health and Human Services, under Contract No. HEW-105-78-1307, Dr. Esther Kresh, Project Officer. Views or conclusions contained herein should not be interpreted as reflecting the official opinion of the sponsoring agency.

AN EVALUATION OF PROJECT DEVELOPMENTAL CONTINUITY
INTERIM REPORT X

ASSESSMENT OF PROGRAM IMPACT THROUGH FIRST GRADE, VOLUME V: IMPACT ON CHILDREN

December 1980

Prepared by:

John Berrueta-Clement
Mary Morris
José Rosario

With the assistance of:

James T. Bond
Gail Pheister
Marjorie Powell
Lynn Spencer
Jana von Fange
Sally Wacker
Peter Williams

Table of Contents

Table of Contents
(continued)

Table of Contents
(continued)

Acknowledgments

Many people have been involved in the development of this report, and in the work leading to the collection and analysis of data reported in this document. We wish to acknowledge the support and assistance of these individuals and groups.

Special thanks go to Dr. Esther Kresh, Administration for Children, Youth and Families. As project officer for the national evaluation of Project Developmental Continuity since it began in 1974-75, she has provided continual support, encouragement, direction, and assistance as needed. Her continuing concerns for the quality of the data and the integrity of the data analysis process have been a constant reminder to us of the potential importance of the findings of this evaluation, in that they may influence future directions of programs at the Administration for Children, Youth and Families.

That influence will be the direct result of the concern for effective programs which has consistently been evidenced by members of the program staff of ACYF. We wish to extend our thanks to Ray Collins, Jenni Klein, Austine Fowler, and Stephen Bedi, who have been supportive of the evaluation effort and interested in the implications of our evaluation results for Project Developmental Continuity and for other ACYF initiatives.

Special thanks go to the coordinators of the PDC sites for their invaluable assistance with the myriad necessary data collection activities. We extend our thanks to those individuals who were coordinators at the time of the grade 1 data collection (1979) and to those individuals who are coordinators at this time: Jesse Beard, Stephen Bedi, Tony Bozich, Nazario Carrillo, Glenda Dodd, Deloris Johnson, Beatrice Kenney, Sande Kirby, Patricia Lanier, Mary D. Levermann, Betty Minor, Geraldine Sanders, Fannie Smith.

Perhaps the greatest thanks should be extended to those many individuals who must remain anonymous to protect their privacy. While we have tried to express our appreciation individually as we work with these people, we also wish to express our appreciation publicly. We therefore thank the students, teachers, parents, district and school administrators, and other individuals who have completed our interviews, taken our tests, and allowed us to observe their classroom behaviors. Without their assistance during the grade 1 data collection and in subsequent years, there would be no evaluation.

Working with the students, teachers, and parents a dedicated team of local data collectors, testers, interviewers, and observers has diligently tracked down students, arranged observation and interview schedules with teachers, located parents, and scheduled (and re-scheduled) interviews as necessary. The national evaluation of PDC has depended upon the energies and professional skills of these individual consultants who mastered the

data collection procedures and then applied those procedures in the field to gather all of the information upon which this evaluation report is based. Many of these individuals, trained during the first year of data collection in 1975, have continued to work with us over the years. Their long-term efforts have enhanced the quality of this evaluation.

Members of the PDC Advisory Panel have provided continual expert review of the evaluation work and valuable suggestions for resolution of technical difficulties. They have been especially helpful in pinpointing various implications of the research findings. For their willingness to consult with us and for their advice about the directions that the evaluation should take, we acknowledge the contributions of Dr. Eva Baker, Dr. Charles Billings, Dr. Jere Brophy, Dr. Robert Dixon, Dr. J. Ward Keesling, and Dr. Luis Laosa.

Within the High/Scope Educational Research Foundation, many staff members have participated in the work which has cumulated in this report. While a few individuals receive recognition as authors, many others deserve recognition as well. Among these are members of our data processing unit: Barbara Bruemmer, Ann Hale, Kim Marker, Jeffrey Moore, Kelly Naylor, Nancy Naylor, Jane Oden. These individuals are responsible for the careful checking in, coding, data entry, data verification, and initial analysis of information collected from a dozen sites, on hundreds of children, parents, teachers, and school administrators. Their attention to detail, their willingness to document their work, their concern with the protection of the privacy of individual respondents, and their flexibility in working with a variety of data collection instruments have all resulted in our confidence in the high quality of the data on which this evaluation report is based.

Another major unit responsible for quality of data is the field operations unit, supervised by Mary Morris. Her calm handling of the many problems which occur during data collection in a dozen sites across the country and her concern for quality in the selection, training and supervision of data collectors, have resulted in a smooth data collection operation. Mary has been ably assisted by Barbara Bruemmer.

Another major team within the Foundation which has been essential to the smooth operation of this evaluation effort is the administrative team. Lynn Spencer has proved invaluable in coordinating activities, gently reminding us of tasks to be done, resolving problems, and handling a million details. Lynn has been supported in this effort by Jana von Fange, who, in addition to her many other responsibilities, has supervised the typing and final proofing of this report. For secretarial assistance in the preparation and production of this report we extend our appreciation to Gail Pheister and Shirley Barnes. Editorial assistance through all the stages of production has been provided by Lynn Spencer: once again we extend our appreciation to Lynn.

A number of other individuals within the Foundation have contributed to this report, and to the evaluation of which this report is one product. David Weikart and Terry Bond have ably served as project monitors at various times in this evaluation effort. Robert Halpern, Art Granville and Allen Smith have completed specific professional tasks as part of this evaluation effort, such as development of data collection procedures and analysis of portions of the data. John Love, project director through the first several years of the PDC evaluation, was involved in all phases of the research and report writing.

To these individuals, named and un-named, we extend our appreciation for their involvement in this work and their continuing interest in the impact of Project Developmental Continuity upon the school districts, teachers and classrooms, parents, and children involved in this major project funded by the Administration for Children, Youth and Families.

Marjorie Powell
Project Director
PDC Evaluation

José Rosario
Interim Project Director

John Berreuta-Clement
Associate Director
Quantitative Analysis

Sally Wacker
Associate Director
Qualitative Analysis

INTRODUCTION

Project Developmental Continuity (PDC) was begun in 1974 by the Administration for Children, Youth and Families (ACYF) as the first large-scale demonstration of coordinated programming between Head Start centers and public schools in 15 communities throughout the country allocated by HEW regions and the Indian and Migrant Program Division. It is hoped that the single most important effect of this undertaking will be to enhance the social competence of the children served--that is, to increase their everyday effectiveness in dealing with their environment (at school, at home, in the community, and in society). PDC also aims to bring about broader and more intensive involvement of parents and teachers in the governance of school affairs.

As part of the overall Head Start improvement and innovation effort, PDC emphasizes the involvement of administrators, classroom staff, and parents in formulating educational goals and developing a comprehensive curriculum. The object is to ensure that children receive continuous individualized attention as they progress from Head Start through the early primary grades. If the program is successful, existing discontinuities between Head Start and elementary school experiences will be reduced by PDC mechanisms that encourage communicating and mutual decision-making among preschool and elementary school teachers, administrators, and parents.

School organizations at the 15 sites received funding to design and implement seven prescribed components:

- Administration: administrative coordination between and within Head Start and elementary school;
- Education: coordination of curriculum approaches and educational goals;
- Training: preservice and inservice teacher, staff and parent training in program-related areas;
- Developmental support services: comprehensive services (medical, nutritional, and social) to children and families;
- Parent involvement: parent participation in policy-making, home-school activities, and classroom visits or volunteering;
- Services for the handicapped: services for handicapped children and children with learning disabilities;
- Bilingual/bicultural and multicultural education: programs for bilingual/bicultural or multicultural children.

At the same time that projects were instituted, the High/Scope Educational Research Foundation was awarded the evaluation contract; the major purpose of which was to provide ACYF with information that would assist it in its efforts to design effective programs for children. The contract called for the collection and analysis of process and impact data involving both quantitative and qualitative methodologies.

The evaluation has proceeded in two phases. From 1974 to 1978 evaluation activities were aimed at analyzing program implementation and assessing the feasibility of doing a five-year longitudinal study that would follow one cohort of children from the time they entered Head Start until they completed third grade.[1] After judging the study feasible, ACYF funded the current phase of the evaluation (1979-1982) to examine the impact of PDC on participating institutions, teachers and classrooms, parents and children in eleven of the twelve sites still participating in the project.

This volume, Impact on Children, is the fifth of a series reporting impact findings as of spring of the study cohort children's first-grade year (1979). Other volumes in the series include:

- Volume I, Assessment of Program Impact Through First Grade: The Context, Conceptual Approaches and Methods of the PDC Evaluation. Serves as an introduction, providing a detailed description of the PDC program and the purposes, methods and guiding framework of the impact evaluation.

- Volume II, Impact on Institutions. Describes findings dealing specifically with PDC's impact on the institutional policies and procedures of participating Head Start centers and elementary schools. These findings are presented in the context of the varied social educational settings surrounding PDC.

- Volume III, Impact on Parents. Investigates the impact of PDC on the parents of children in the evaluation cohort and, in a preliminary fashion, the relationship between family characteristics and outcome variables.

- Volume IV, Impact on Teachers. Reports impact findings on teachers and classrooms. These impacts reflect treatment-related outcomes as well as outcomes regardless of treatment.

[1]The results of this phase of the evaluation are described in: Love, Granville and Smith, 1978; and, Smith, Love, Morris, Spencer, Ispa and Rosario, 1977.

- Volume VI, Summary of Impact on Institutions, Teachers and Classrooms, Parents and Children. Summarizes the evaluation results for 1979, when the cohort of children being studied in the evaluation had completed grade 1. Results are presented for each of the four major areas: institutional policies and procedures, teacher attitudes and behaviors in the classroom and with parents, parent attitudes and behaviors in relation to their child's school, and the achievement of children. In addition, the volume summarizes the initial analyses of interrelationships between the four major areas, such as the relationship between teacher attitudes and parent behaviors concerning involvement with their child's school.

This volume reports on the impact of the PDC program on the evaluation study's cohort of children. It is organized into five major sections, plus a summary and two technical appendices. This chapter presents a brief history of the PDC program and its evaluation.

Chapter II describes the conceptual framework guiding the study of PDC processes and effects on children. This framework has made it possible for us to begin to "model" the concept of Project Developmental Continuity as well as the kinds and directions of changes necessary for its institutionalization. It is presented as two different "models": a conceptual model that describes ideally the intended effects of PDC, and an analytic model that describes operationally the change flow expected and required to bring about the intended effects. Chapter II also describes the outcome measures that are used to assess the children's performance, and how these measures fit into the conceptual and analytic models.

The data collection and analysis procedures required by a study of this magnitude and complexity are discussed in Chapter III under the general title of "Methods." Chapter IV presents the descriptive findings regarding the sample and the characteristics of the instruments. Chapter V describes the results of the child outcome measures analyses. In Chapter VI, we summarize the major findings and discuss implications for the future.

II

A FRAMEWORK FOR STUDYING PDC PROCESSES AND EFFECTS ON CHILDREN

The evaluation has been largely shaped by a particular conception, derived from the PDC guidelines, of the intended effects of PDC and the sequence of changes expected and required to bring about those effects. Before describing the design and methodology of the evaluation, we will in this section attempt to make this conceptual framework more explicit. This discussion has three parts. In the first two, we present a general model of the intended effects of PDC, along with a consideration of the PDC "treatment" and how, as described in the guidelines, it was intended to produce the desired effects. In the third part we describe the process that was used to move from the basic framework to the specification of particular variables and appropriate data collection instruments for this phase of the evaluation.

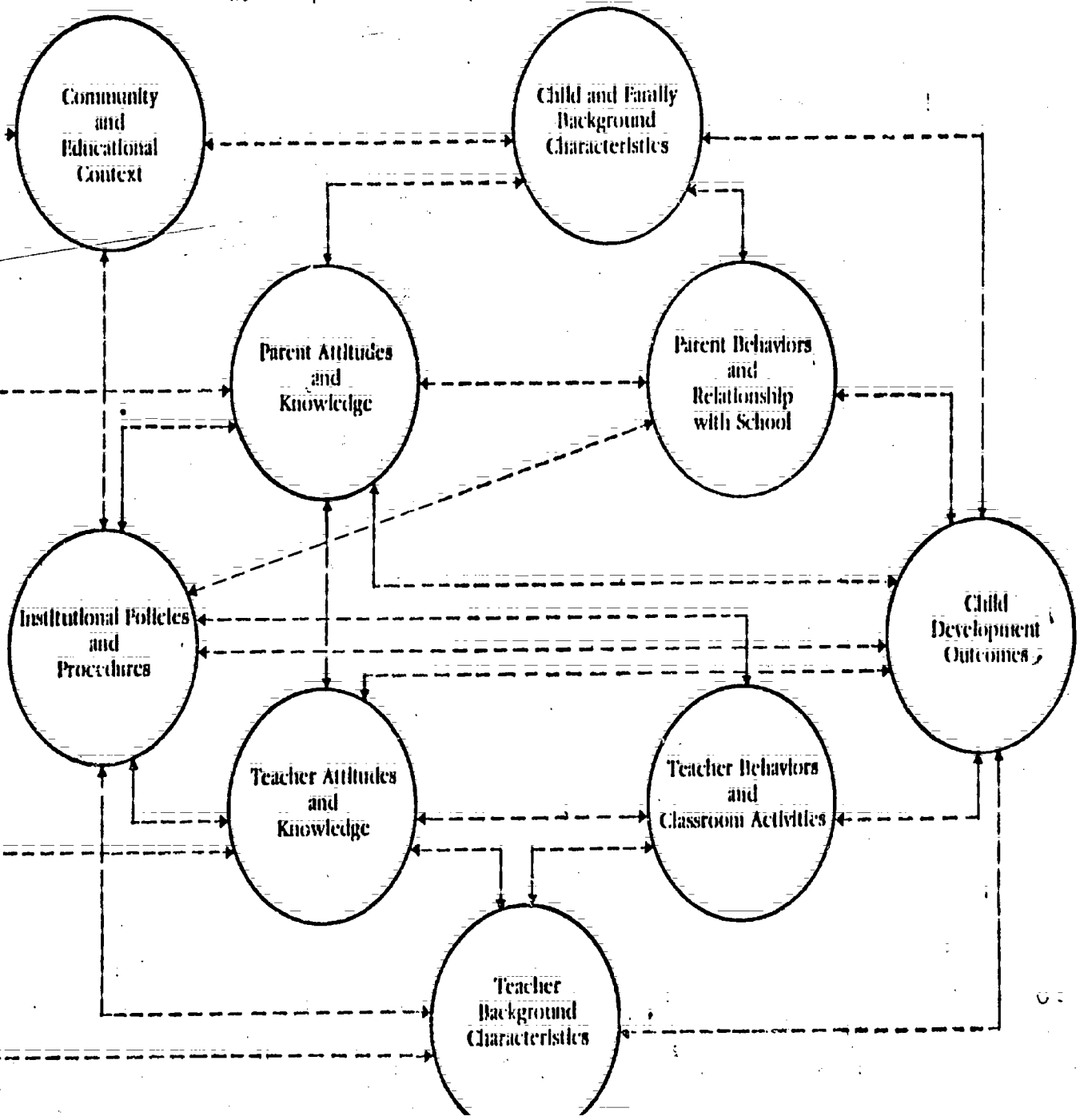
Some Orienting Assumptions: The Concept of Developmental Continuity

The basic assumption underlying the PDC program and consequently this evaluation is that the condition of developmental continuity implies a complex interaction involving an array of factors, both within and outside the school. As a result of this assumption, PDC was designed to be a comprehensive intervention into many aspects of the school, home and community. However, although the implications of this basic assumption pervade the program, the PDC guidelines never fully explicate this assumption.

In order to design an evaluation that is sensitive to the particular goals of the PDC program it was necessary to distill from the guidelines the concept of developmental continuity that appears to have shaped program guidelines. Figure 1 summarizes the results of this exercise. We must emphasize that this conceptualization is not at present a theory to be tested by the data. Rather, it represents an orienting framework that has provided a basis for generating an analytic model, out of which have come research questions, variables, and data collection methodologies. We have used this orienting framework to guide the analysis and reporting of evaluation data.

Simply stated, the conception of developmental continuity implicit in PDC suggests an interactional model that appears to include: (a) a child's intellectual, social, and physical development and background and experiences in home and school; (b) the attitudes, knowledge and background characteristics of parents and teachers; (c) the policies and procedures that prevail in the public school or Head Start center; and, (d) the broader political, social and economic context of the school district and community.

Figure 1
The Conception of Developmental Continuity Assumed in PDC



We will return later to consideration of how each of the classes of factors in Figure 1 was defined operationally for this evaluation, and of what variables were measured in each domain. For the moment, however, the following general definitions will suffice:

- Child development outcomes. These, of course, are the ultimate concern of the PDC program. The stated goal of PDC is to enhance children's "social competency." According to the guidelines, social competence includes intellectual achievement, health and nutrition, social-emotional and language development, physical and mental health, and learning attitudes.

- Parent behaviors. This domain includes parent behaviors toward the child in the home, and the role that the parent plays in school life.

- Parent attitudes and knowledge. Especially important in this domain are parent attitudes toward the school or center and parent knowledge of child development and available community resources.

- Teacher behaviors and classroom activities. This domain refers to the child's experiences in the classroom and to the role of the teacher in these experiences. It includes the physical environment that the teacher creates for the child in the classroom, the instructional approach that the teacher employs, the management style of the teacher in his/her dealings with the class, and the general climate that the teacher establishes in the classroom for the children.

- Teacher attitudes. A broad and often-noted domain in the program guidelines, this category refers to teachers' instructional practices and their perceptions of, and attitudes toward parents, particularly parent involvement in their classrooms, and their personal educational philosophy.

- Institutional policies and procedures. This domain includes the activities and procedures that are found outside the classroom, but which influence what goes on in the classroom. Such policies and procedures include the decision-making bodies and mechanisms that exist in the school, the management structure found in the school, procedures for providing services to children either inside or outside the classroom, patterns of communication and coordination in the school and between the school and other institutions, and training that the school provides for teachers, parents, and staff.

- Community and educational context. No school or family exists in a vacuum. The program guidelines recognize that everything that occurs in either setting is shaped and on occasion constrained by cultural, political, and economic factors in the community, and by priorities, policies, and programs of the school district. Another important feature of the community context is the services for families and children that are available from agencies outside the school.

- Child and family background. Although not generally susceptible to change by school programs, the background of the child and his or her family are recognized in the guidelines to be important determinants of development. This domain includes such factors as ethnicity, SES, parents' education and employment status, language spoken in the home, and prior preschool experience.

- Teacher background characteristics. The guidelines say little about particular effects of specific background characteristics, but they and the literature do suggest that such factors are important influences on the teachers' behavior and ultimately on child development. The guidelines refer specifically to certain experiences that at least some program teachers should have had, such as training in bilingual education, or training in child development; the literature also suggests that ethnicity, number of years of teaching experience, and experience in special projects also influence teachers' professional behavior.

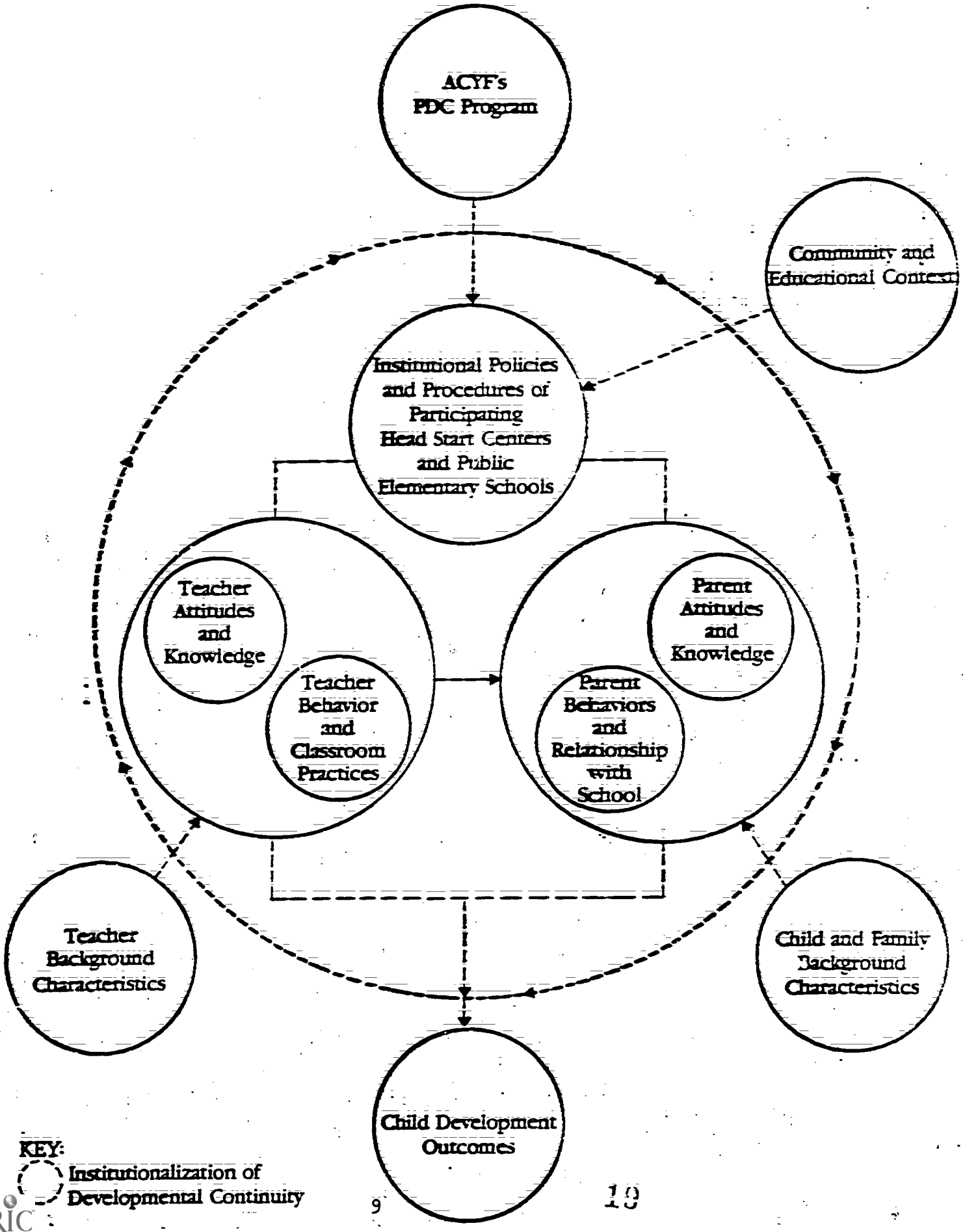
The PDC guidelines do not discuss the precise interactions that are assumed to exist among these various factors. Consequently, Figure 1 portrays only a cycle of continuous interactions that is driven by incremental changes acting on each other in a positive way. One objective of this evaluation will be to explore and describe the strength and direction of relationships between variables within each domain.

However, the guidelines are quite clear in specifying an order in which changes occur to produce impacts on elements of the interactive cycle represented in Figure 1. Any program that seeks to create developmental continuity must first impact on institutions, and through them on parents and teachers, before it impacts on children. Figure 2 presents an analytic model that describes the direction of this change flow.

As shown, PDC is expected to produce first certain interactive conditions favorable to the institutionalization of developmental continuity which are then expected to lead to changes in child development outcomes. The operational strategy for producing these favorable conditions is to bring about the institutional or structural changes that then make it possible for institutional actors (administrators, teachers and parents) to engage in educational practices that are mutually reinforcing and developmentally continuous. At first, it is expected that the change flow will be moderated by the community and educational context as well as teacher, child and family background characteristics. But ideally, of course, the expectation is to create a chain of interactive changes that spread over time to eventually produce the kind of developmental cycle illustrated in Figure 1. In a sense, then, the analytic model of Figure 2 represents an early stage in the PDC implementation process, and the ultimate steady state is represented by Figure 1.

Figure 2

The Change Flow Assumed in PDC



What Is the PDC Treatment?

We have noted that the ultimate goal for the PDC program is to enhance the social competence of the children it serves by providing developmental continuity. Some of the assumptions implicit in the guidelines about the interactive factors involved in this process have already been examined. The question we must ask next is exactly how the PDC project was intended to impact upon the factors that the guidelines assume will be present in developmental continuity. In other words, what is the PDC treatment?

Again, the program guidelines offer the best starting point for answering this question. In the introduction to these guidelines the following statement appears:

> "Project Developmental Continuity is aimed at promoting greater continuity of education and comprehensive child development services for children as they make the transition from preschool to school...Developmental Continuity, as it is used here, can be defined as planned programs, structures, systems, or procedures by which adults provide children with experiences that foster and support continuous development." (emphasis added)

Project Developmental Continuity seeks to enhance children's social competency by creating greater continuity among children's experiences in the school and between children's home and school experiences. The guidelines do not attempt to specify what continuity of experience should look like, but instead outline a set of planned programs, structures, systems, or procedures that, if implemented, will result in the desired continuity. These structures, then, are the basic PDC treatment that should be present at all sites; within this general framework each site is free to develop its own program.

Table 1 contains brief descriptions of the structures or programs prescribed in the guidelines for project sites. These prescriptions outline a set of activities for all PDC programs to implement. Following the earlier model, these guidelines are aimed at the classroom, at parents, and at the school or center as an institution.

Identifying an Evaluation Methodology Appropriate for the PDC Treatment

Having specified the PDC treatment as described in the guidelines, the next step was to develop an evaluation design that was appropriate to the goals of the PDC program. Although this process also began with the program guidelines it was necessarily shaped by other considerations

Table 1

The PDC Treatment as Described in the Guidelines

Planned Programs, Structures, Systems or Procedures that Foster and Support Continuous Development

At the Institutional Level

Planning and Decision Making

1. Formalized broad representation in decision-making groups including parents, staff (Head Start and elementary), community representatives involved in education, health, nutrition, and social services.
2. Procedures for ongoing discussion and refinement of the curriculum that include parents, teachers, aides, etc.
3. Establishment of a formal or informal internal assessment system for monitoring the school's progress toward meeting its goals and objectives.

Management

1. Assign responsibility for education, handicapped, bilingual, etc. to specific individuals at Head Start and elementary levels.
2. Provisions for coordination from Head Start through grade 3 of services to meet the educational and social needs of handicapped and bilingual children.
3. A coordinated parent involvement program from Head Start through grade 3.

Training

1. Provide training on decision making and policy making for members of decision-making groups.
2. Provide training on the goals and objectives of both the Head Start and elementary programs.
3. Provide training to make staff and volunteers sensitive to special needs of handicapped children.
4. Provide training for parents in how to work with teaching and administrative staff.
5. Provide training for classroom volunteers.
6. Provide training for parents in how to work with their own children.
7. Provide training for parents in child growth and development.

Table 1
(continued)

Training (continued)

8. Provide training for parents in available community resources.
9. Provide training for teaching staff in meeting the needs of bilingual children.
10. Provide training for teaching staff in the principles of first aid, health, and safety practices.

Communication and Coordination

1. Communication between decision-making bodies and Head Start and elementary school parents.
2. Regularly scheduled communication and coordination between Head Start and elementary teaching staff.
3. Continuity of record-keeping, Head Start through grade 3.

Provision of Services

1. Provision of a broad range of medical, dental, mental health, and nutrition services.
2. Comprehensive screening and diagnostic assessment of every child upon enrollment.
3. An annual survey to identify handicapped children.
4. Provision of an interpreter when needed.

At the Level of Classroom Activities

A Continuous Coordinated Curriculum

1. Develop or adopt a compatible, coordinated curriculum from Head Start through third grade.
2. Have a curriculum that facilitates the learning of basic educational skills for reading, writing, and computation.
3. Have a curriculum that provides continuity of educational and developmental experiences, Head Start through grade 3.
4. Develop a curriculum plan that includes goals and objectives statements in each subject or developmental area.

Table 1
(continued)

Individualized Instruction

1. Curriculum must be developmentally appropriate.
2. Instruction must be individualized.
3. Develop a diagnostic and evaluative system that enables teacher to pinpoint developmental levels of each child based on the child's diagnosed strengths and weaknesses.
4. Former teachers consulted when planning educational objectives.

Multicultural Perspectives

1. Provide bilingual/multicultural classroom activities, materials and resource persons for all children.
2. Develop a compatible Head Start-elementary school approach regarding bilingual education.

Classroom Services for Handicapped Children

1. Handicapped children mainstreamed to the maximum extent possible.
2. Early diagnosis and evaluation of children with learning disabilities.
3. Special materials, structural changes, or classroom reorganization provided as appropriate for accommodating handicapped children.

Whole-Child Perspective

1. Have a curriculum that encourages the physical and social-emotional growth of children.
2. Health education and nutrition integrated with other educational objectives and activities.
3. Meals and snacks used as an opportunity for learning.
4. Provide nutritional services that reinforce good aspects of foods served at home.
5. Familiarize children with health services they will receive prior to delivery.

Use of Community Resources

1. Bilingual/multicultural resource persons used in the classroom.

Table 1
(continued)

At the Level of the Home and Home-School Activities

Home-School Communication

1. Parents involved in planning educational objectives for their children.
2. Parents given summary of records on health, medical services and immunization.
3. Parents familiarized with available health services.

Parent Involvement in School Life

1. Parents involved in all decision-making bodies.
2. Parents involved in all school decisions.
3. Activities provided for parents that relate to cultural dynamics.
4. Parents used as resource persons in the classroom.
5. Parents involved in classroom activities, special parent events, activities that stress home-school continuity.
6. Parents involved as observers, aides or volunteers in the classroom.

Home Activities with Children

1. Parents encouraged to become involved in health care process.

as well. First, PDC is not a static program, launched and maintained by an immutable set of guidelines. Local programs through their experiences and interactions with national ACYF staff have created altered perceptions of what PDC is and should be. These altered perceptions had to be accommodated in the evaluation design. Second, the PDC evaluation itself exists within a broader research and policy environment. New issues and questions are emerging regularly that could appropriately be addressed in the PDC evaluation without compromising the basic evaluation objectives. Consequently, certain research questions and variables have been added to the study in response to ACYF information needs that are not necessarily unique or even directly tied to the PDC treatment as defined in the guidelines. Finally, there are many audiences for the PDC evaluation, each with its own information needs. These audiences include policy makers in Washington, the research and evaluation community, and of course practitioners in the field. Insofar as possible, the needs of these audiences have been accommodated within the evaluation design.

Before outlining the research questions and associated variables for the evaluation, a few words are in order about the process that was used to develop the study. The RFP for the second phase of the evaluation specified that the contractor was to examine the impacts of the PDC program on children, on parents, on teachers, and on the schools and centers as institutions. The RFP also specified that these impacts were to be assessed using a variety of structured and unstructured methodologies, from classroom observations to interviews and document analysis.

Early in the contract, several representatives from the various constituencies of the PDC program were invited to High/Scope's Ypsilanti, Michigan headquarters to "brainstorm" about the PDC treatments and the impacts that could plausibly be expected in each impact domain. This panel included a coordinator from the PDC project in West Virginia, a technical assistance consultant familiar with several sites, and a former ACYF project officer familiar with ACYF's policies. The panel met with High/Scope staff for three days and produced a long list of (a) plausible impacts and (b) variables that might be measured to assess these impacts.

This initial and admittedly massive list of impacts was next sorted, pruned, refined, and revised by project staff and presented to the PDC Advisory Panel in October 1978. Breaking into work groups that concentrated on each impact domain, panel members worked with project staff to further prune the list and to establish priorities among the many variables that might be assessed in each area. This refined list became the basis for all instrument development. Further modifications and refinements have been made to this basic list as new information needs have been identified through ongoing interactions with PDC program staff at ACYF.

Research Questions, Constructs, and Variables

This phase of the PDC evaluation is designed to address three basic questions:

1. *What impact has the PDC program had on (a) children's development, (b) parents' knowledge and attitudes, (c) parents' behaviors, (d) teachers' attitudes and knowledge, (e) teachers' behavior and classroom activities, and (f) institutional policies and procedures?*

2. *Irrespective of treatment, what factors or patterns of factors help account for meaningful outcomes in each domain?*

3. *To what extent do these factors affect the relationship between the PDC program and its impacts?*

Stated differently, the first task of the PDC evaluation is to determine PDC program effects through comparisons of PDC and comparison teachers, parents, and children on selected variables. For example, the frequency of parent visits to PDC and comparison schools is compared to determine whether PDC has had any impact on that aspect of parent involvement in schools. The next task is to explain the results of these comparisons using whatever qualitative and quantitative information is available. For example, at sites where there are relatively few or no differences between PDC and comparison parents' involvement in the school, we may find that the comparison schools have instituted a parent involvement program patterned after PDC's. It might be reasonable to conclude from this that, contrary to appearances, PDC has indeed had an impact upon parent involvement in the schools in question, and that impact has diffused to the comparison institutions.

Having examined the similarities and differences between PDC and comparison groups along various dimensions, the final task for the evaluation is to examine the relationships among child, parent, teacher, institutional, and community variables, disregarding the PDC/comparison grouping. Extending the preceding example, we might discover that schools with active and successful parent involvement programs, be they PDC or comparison, tend to have similar institutional policies or procedures (such as regular newsletters, parent training programs, and designated parent involvement coordinators) that foster greater involvement by parents in school activities. While findings such as these may not reflect directly on the effectiveness of the PDC treatment, they would be of obvious interest to educators and policy makers wishing to expand the role of parents in school programs.

Constructs Addressed by the Evaluation

As we have said, a pervading concern in the design of this evaluation has been ensuring that the domains and variables measured are indeed relevant and appropriate to the objectives of the PDC program. The development process that was followed to accomplish this end has already been described. Following this process a set of constructs was identified in each impact domain for attention by the evaluation. These constructs are listed in Table 2.

For the most part, these constructs follow the conceptualization of the PDC treatment that was mapped in the program guidelines and refined by ACYF and project staffs (see Table 2). Thus, the constructs described in the table generally represent the areas in which PDC was supposed to have impacts, and areas in which the nature and direction of PDC/comparison differences could be predicted. There are some exceptions to this general rule, however. Most exceptions are found in the domain of Teacher Behaviors and Classroom Activities, where several constructs—Structure and Content of Classroom Environment, Classroom Climate, Intellectual Stimulation, Classroom Management, and Instructional Approach—were added despite the fact that the guidelines are virtually silent about the specific impacts that PDC should have in these areas. They were included in the evaluation because other research has indicated that behaviors in each may contribute significantly to child development outcomes. Although few hypotheses could be formulated about PDC/comparison differences in these areas, they were nonetheless included because of their potential utility in answering Research Questions 2 and 3.

Variables and Data Sources

For each construct in every domain an array of variables was identified through consultation with ACYF, local project staff, and outside experts, following the procedures outlined earlier. For each variable, decisions were made about the best sources of information and data collection methodology. Wherever possible an attempt was made to "triangulate" on the desired information by collecting data on the same phenomenon in multiple ways from different sources. Table 3 lists the data collection instruments and methods developed for the evaluation; more extensive descriptions of the instruments can be found in Volumes II, III, IV, and V of the series. The appendix to Volume I contains a list of the variables addressed by the evaluation, the sources for information on each variable, and the hypothesized directions of treatment effects.

Table 2

Domains and Constructs Addressed by the PDC Evaluation

Child Development Outcomes

- Academic skills and abilities
- Health and nutrition status
- Social-emotional development
- Learning attitudes
- Classroom behavior

Parents' Behaviors

- Role of parents in school life
- Parent-child activities in the home

Parents' Knowledge and Attitudes

- Parents' attitudes toward the school as an institution
- Parents' perceptions of the schools' help in meeting the needs of their families

Teachers' Behaviors and Classroom Activities

- Structure and content of classroom environment
- Delivery of special services to children
- Classroom climate
- Meeting needs of handicapped children
- Intellectual stimulation
- Home-school continuity
- Contacts with other teachers
- Instructional approach
- Classroom management
- Individualization of instruction
- Use of community resources
- Meeting affective/emotional needs
- Multicultural perspective

Teachers' Attitudes

- Attitudes toward parental involvement
- Perceptions of change
- Attitudes toward the school/center

Institutional Policies and Procedures

- Planning and decision making
- Provision of services
- Use of community resources
- Communication and coordination
- Training

Table 3

Data Collection Methodologies*

Child Development Outcomes

Instrument	Type	Abbreviation
Peabody Individual Achievement Test	Individually administered published test	PIAT
McCarthy Scales of Children's Abilities	Individually administered published test	MSCA
Bilingual Syntax Measure	Individually administered published test	BSM
Preschool Interpersonal Problem Solving Test	Individually administered published test	PIPS
Child Interview	Semistructured interview followed by interviewer ratings	CI
Child Rating Scale	Teacher ratings of individual children	CRS
Pupil Observation Checklist	Tester ratings of child's behavior during test administration	POCL

Parents' Attitudes, Knowledge, and Behaviors

Parent Interview	Structured interview with parents of children in test cohort	PI

*See Appendix A for complete descriptions of instruments.

(continued)

Table 3
(continued)

Teachers' Activities, Knowledge, and Behaviors

Instrument	Type	Abbreviation
Teacher Interview	Structured interview	TI
Classroom Environment Observation	Checklist and rating form	CEO
Classroom Activities Record	Time-sampling observation and rating form	CAR
Focused Observations	Semistructured observations and rating form	FO

Institutional Policies and Procedures

Administrator Interview	Structured interview	AI
Case Studies	Documents prepared by Pacific Consultants for ACYF in 1978-79	—
Site Visits	One-week visits by High/Scope staff	—
Site Records	Minutes, training records, etc. kept by local project staff	—

How Child Development Outcomes Fit Into the Conceptual and Analytical Models

This section describes how assessments of child outcomes are tied into the conceptual and analytic models presented earlier. We begin by reviewing the specific child development outcomes of interest in the PDC evaluation, and then discuss the relationship between these outcomes and other variable domains (such as parent and family background, teacher background and teacher and parent outcomes) within the analytic and conceptual models of the PDC process.

Child development outcomes are grouped into five constructs: academic skills and abilities, health and nutrition status, social-emotional development, learning attitudes, and classroom behavior. In the grade one assessment two constructs (health and nutrition status and classroom behavior) did not receive major attention; this situation will change in succeeding evaluation years, as technical issues related to consideration of variables in these constructs are resolved. Consequently, the measures of child development outcomes upon which analytical emphasis has been placed are those in the remaining three constructs. Table 3 lists the child development measures used in the PDC evaluation; these can be grouped into the three constructs in the following way:

- Academic skills and abilities: PIAT-Reading and PIAT-Math; Bilingual Syntax Measure; McCarthy Scales; and Verbal Fluency subscale.

- Social-emotional development: Preschool Interpersonal Problem-Solving Test; POCL-2 "Task Orientation" subscale; CRS-1 "Self-Assurance" subscale; CRS-2 "Aggressiveness" subscale; CRS-3 "Dependence" subscale; and CI-1 "Attitude Toward the Teacher" subscale.

- Learning attitudes: POCL-1 "Task Orientation" subscale; CRS-4 "Academic Motivation" subscale; and CI-2 "Interest in Reading" subscale.

The measures, including their administration and scoring procedures, are described in Appendix A of this volume. For analytic purposes in the present report they have been treated singly or taken together as a whole. Future interim reports will deal with the integration of child outcome measures into major research constructs.

Child development outcomes, in the context of the model described earlier, are seen as dependent on a complex of constructs from several domains. Constructs in the teacher and parent domains are seen as those most closely linked to child-level constructs. Specifically, certain teacher behaviors and classroom practices, on the one hand, and certain parent behaviors toward the child and parental relationships with the

schools, on the other, are viewed as associated most directly with changes in child outcomes. In turn, these teacher constructs are deemed in the model to be associated with teacher attitudes and knowledge, and ultimately with teacher background characteristics. The parent behaviors are associated, for their part, with parent attitudes and ultimately with parent and family background characteristics. The model depicted in Figure 2 can be extended even further back toward institutional features; but such linkages are not explored here.

The spring 1979 data collection (grade one for the study cohort) is the first occasion on which data were collected simultaneously in the parent, teacher, classroom and institutional domains. In this volume, we have summarized the results of the first limited explorations of association between outcomes for children and variables in teacher and parent domains. The results of the study for child outcomes are explored through three major questions:

1. What is the impact of PDC on child outcomes at the end of grade one?

2. What other variables (in domains specified by the model of Figure 2 as most closely linked to child outcomes) contribute to explaining grade one child outcome findings?

3. To what extent do these other variables alter the nature of the educational treatment's impact?

In responding to the second and third research questions, the variables selected for consideration come from three construct areas:

- teacher background characteristics
- teacher "outcomes": teacher behaviors and attitudes that might legitimately be considered as resulting from processed in the PDC program
- family background characteristics.

In Chapter V, we present analyses of the association between these variables and child outcome measures. We also discuss in greater detail some of the issues raised in these analyses and in the interpretation of their results.

III

METHODS

Data Collection Procedures

To establish a data collection routine that would result in data of the highest possible quality, the procedures followed in the preceding data collection periods were continued, with minor modifications:

- An organizational structure for individuals involved in the data collection effort was outlined, role responsibilities were defined, and detailed training manuals were produced.
- Training models were designed that specified tester performance standards and provided for sessions with large-group, small-group and individualized instruction, daily reviews of each field staff's performance, and discussion of potential problems.
- Onsite monitoring of field staff by trainers was conducted prior to the start of the actual data collection.
- During the data collection period, testers were responsible for monitoring each other's performance on a weekly basis.
- Site coordinators collected completed data each week and checked it for obvious errors or omissions before sending it to the High/Scope Foundation.

Each of these procedures is discussed below.

Field Organization

Job announcements for tester positions were posted in all sites by the local PDC staff. Applicants were then interviewed by High/Scope staff and final hiring decisions were based on their experience in working with children as well as their performance on a mock test or interview and their perceived ability to interact effectively with school staff. The roles of the personnel who conducted field data collection were explicitly defined in the High/Scope PDC Field Procedures Manual in order to clarify and systematize responsibilities. In addition to actual testing, interviewing and observing by field staff, one tester and observer from each site was designated site coordinator. Site coordinators' responsibilities, in part, included informing the site's PDC coordinator about the start of the data collection; setting up and chairing

a meeting with the first grade teachers involved in the evaluation, or contacting them individually; maintaining regular contact with High/Scope's liaison person to monitor the site's data collection effort and to discuss any problems the site was experiencing; and checking the completed data each week before mailing the forms to High/Scope for processing. From start to finish the data collection effort took approximately nine weeks at each site.

Training Procedures

Training sessions for both High/Scope trainers and PDC data collection field staff were held in March 1979 at the High/Scope Conference Center in Clinton, Michigan. Since all five High/Scope tester-trainers had been involved in previous PDC training sessions, a brief, one-day session was scheduled for them during which they reviewed and practiced the child measures and discussed the tester-training agenda and training methods. PDC testers attended a four-day training session Other training sessions attended by all field staff included those on interviewing techniques and field logistics.

Tester training. During the tester training session, each test was presented and practiced in small groups. Practice sessions involved the use of test "scripts" which consisted of test instructions, child responses, and rationales for scoring. In using the scripts, two testers would pair up and one (the "child") would perform as indicated on the script while the other tester administered the test without the script. This provided an excellent learning situation because the child responses included on the script covered all the administration rules and gave the testers a chance to work with and correct each other. Also, since the majority of testers were experienced PDC testers they were able to help the new testers with test administration procedures and give advice on their "tried and proven" techniques for establishing rapport and interacting with children and teachers.

In order to insure that testers administered the tests in a standard manner, each tester was systematically "checked-out" on all of the child measures before the end of the training session. During this procedure, a High/Scope trainer played the role of the child (also recording the "child's" responses) while a tester administered one or more of the child measures to her. The High/Scope trainer (acting as the child) responded in standard ways to each item on each test in order to insure that: (1) each tester was exposed to the same situations, and (2) the trainer could assess the tester's handling of critical child responses. For example, on the PIPS interview, there are specific things for a tester to say if a child gives an unrelated answer, a repeated answer, refuses to answer, and so on. By exhibiting all these behaviors in the check-out situation, trainers were able to assess the tester's understanding and expertise in administering each of the child measures.

Standards were set for acceptable performance during the tester check-outs, and if these standards were not met, additional training and practice was prescribed. Check-outs were then repeated at a later time during the training session to insure correct test administration.

Monitoring

Onsite tester monitoring. Onsite monitoring occurred the week following the training session in all sites where new testers had been hired. During the monitoring session each of the testers administered the PDC measures to a child while a High/Scope trainer observed the interaction. After the session, the High/Scope trainer provided feedback (if necessary) to the tester on ways to improve her interactions with children. This monitoring procedure served two purposes: (1) it gave the trainer an indication of how well the new tester was able to establish rapport and interact with children; and (2) it helped alleviate some of the anxieties the inexperienced testers felt about administering the measures to children.

Weekly tester monitoring. During the course of each testing week, testers at each site alternately monitored each other. One tester acted as monitor and simultaneously completed the test booklets and the individual monitoring forms for each test. After the session, the "monitor" and tester discussed any errors and the monitoring booklets and forms were sent to the supervisor of field operations at the High/Scope Foundation to be reviewed.

Weekly Pre-Transmittal Data Checks

Testers were required to give or send their completed data to their respective site coordinators at the end of each week. These staff then checked the tests for recording/scoring errors. (Site coordinators and testers reviewed a checklist specifying what to look for when reviewing each completed booklet, e.g., "Is the identification complete?" "Did the interviewer skip an item?") Errors were pointed out to the particular tester and, if necessary, further training was provided by the site coordinator. The site coordinators also kept track of all completed data (in addition to the individual records each tester and observer kept) and were responsible for mailing the completed data to the High/Scope Foundation on a weekly basis.

Recording and Scoring of Data

In addition to the site coordinators' pre-submittal check, data collected by the testers and observers were also checked by the supervisor of field operations at the High/Scope Foundation. The supervisor of field operations identified any errors in recording or coding and notified the site coordinators, who then discussed the errors with the testers and observers at the site.

Once the raw data were screened for accuracy at High/Scope, they were sent to the Foundation's data processing section to be tagged with unique identification numbers for each student, scored and verified, and then keypunched and verified.

Data Collection Sequence

Once the sample children for the evaluation were located in the district schools, the field staff divided the classes among themselves. In making these divisions two factors were taken into account: (1) the order in which the classes were to be completed was such that testers would be collecting data simultaneously in the PDC and comparison schools, and (2) each field staff member would be testing or observing in both PDC and comparison observer bias for either group.

Data Analysis Procedures

Chapters IV and V of this report present the results of five sequential stages of analysis of PDC data, focusing on:

- descriptive characteristics of PDC and comparison group samples for which data were collected in spring 1979;
- attrition patterns in the spring 1979 samples, and their consequences;
- characteristics of the instruments in the spring 1979 PDC battery;
- impacts of the PDC program on participating children, both over time and as of spring 1979, when almost all of the children in the study cohort were in grade 1;
- preliminary analyses of factors other than treatment that might help to account for the identified impacts on child outcomes; and
- preliminary analyses of the relation between teacher characteristics and child outcomes independent of treatment.

Descriptive Characteristics of the Samples

In order to understand the composition of the PDC and comparison child samples for which data were collected in spring 1979, descriptive statistics were computed and tabulated for these samples at each site and for all sites combined.

Previous reports have defined an analytic subsample for child-level analyses, a subsample which excluded children with defined handicaps or with a dominant language other than English. The analytic subsample, in other words, has in the past been smaller than the full sample for whom data were collected. In analyses of spring 1979 data, the full sample was employed except for those analyses that include pre-1979 data for which children tested in Spanish at some time points had to be excluded. The rationale for inclusion of handicapped and Spanish-speaking children in the analytic sample for this report is discussed in Chapter IV.

Attrition Patterns

Representativeness of remaining sample children. Children who departed the PDC and comparison group samples through spring 1979 were compared with the children remaining in the study samples on a number of background variables and on fall 1976 test scores. The purpose of these analyses was to determine whether any differences between these two groups of children could be identified; such differences would indicate that selection effects had operated to diminish the representativeness of the samples remaining. The hypothesis of attrition-induced changes in the samples was evaluated by means of univariate and multivariate one-way analyses of variance, and for nominal data by chi-square analyses.

Comparability of remaining samples. The samples of PDC and comparison group children tested in spring 1979 were compared on background variables and entry-level test scores to determine whether these groups might still be considered equivalent in their characteristics at the time of program entry (fall 1976). Again, univariate and multivariate one-way analyses of variance and chi-square tests were employed.

Characteristics of the Instruments

Because the principal instruments assessing child-level outcomes in the spring 1979 testing battery have, in earlier PDC analyses, passed through at least two screenings based on psychometric criteria, no further screening was considered necessary before data collected with these instruments were entered into outcome analyses. Psychometric analyses are presented in this report principally for purposes of further instrument definition and documentation. Psychometric properties reported in Chapter IV for the BSM-English/BSM-Spanish, PIAT-Math, PIAT-Reading, Child Interview Scales, Child Rating Scales, POCL 1 and 2, and the PIPS include:

- overall central tendency, dispersion and distributional characteristics of summary scores or scale scores (usually as means, standard deviations and histograms)

- assessments of reliability (as internal consistency estimates)
- assessments of validity
- assessments of stability
- assessments of sensitivity to change
- variable intercorrelations
- factor structures
- relation to an assessment of "social competence"
- item analyses

Analytic Strategies for Examining the Impact of PDC

Evidence of PDC's influence on children is examined in the first part of Chapter V. The analytic strategies used there to measure PDC's effect are described briefly below. A number of analytic questions are posed:

1. Is there a difference between the PDC and comparison groups for each of the child outcome measures on which data were collected in spring 1979?

2. Is there a difference between the PDC and comparison groups in outcome measure profiles, when all child impact measures obtained in spring 1979 are combined?

3. Is there a difference in the growth curves of the PDC and comparison groups over time, considering each outcome measure separately?

4. Has a difference between the PDC and comparison group children emerged between assessments in the spring of their kindergarten year and spring of their grade 1 year, considering each outcome measure separately?

These questions are addressed through univariate and multivariate analyses of covariance. The first analytic question is addressed through a series of univariate analyses of covariance; the second, through a single cross-sectional multivariate analysis of covariance incorporating all child-outcome measures obtained in spring 1979 as dependent variables. The third and fourth questions are addressed together through the use of multivariate repeated measures analyses examining one outcome instrument at a time, including tests for growth trends and for change patterns over

the period subsequent to earlier measurements. Covariance adjustments of the data are performed in order to compensate for differences between groups at entry, and to adjust for differences between treatment groups associated with sex, ethnicity and prior preschool experience.

Analytic Strategies for Assessing Factors Affecting PDC Impact

The effects of site and the interactions of site with treatment in relation to child outcomes were analyzed in the course of answering questions relating to program impact. The influence of these factors on child outcomes is discussed in Chapter V. Also discussed in Chapter V are the effects of the covariates used in the child impacts analyses: sex, prior preschool experience, and ethnicity.

The effects of language spoken at time of entry into the study, age at entry, and two family characteristics (mother's education and number of siblings in the family), are also explored in Chapter V. The analytic approach used in these preliminary explorations is multiple regression. The dependent variables are those child outcome measures that show evidence of outcome differences in the assessment of program impact. In order to establish whether the effects of these variables modify statements about the impact of the program, regressions are conducted in two ways: both including and excluding a dummy variable for treatment.

Analytic Strategies for Assessing the Relation Between Teacher Characteristics and Child Outcomes

The relation between teacher background characteristics, teacher attitudes and behaviors, and child outcomes is explored in a preliminary fashion in Chapter V. Technical problems in the consideration of relationships between other variable domains (such as the parent domain) and child outcomes are also discussed in that chapter.

Again, multiple regression is the analytic approach used, with the dependent variables being various measures of child outcome.

IV

SAMPLE AND INSTRUMENT CHARACTERISTICS

Characteristics of the Sample

General Description

A total of 551 children were tested at 11 PDC sites in spring 1979. Table 4 displays the numbers of children in the PDC and comparison groups at each site and describes the composition of each group.

All of the children tested entered into the analyses presented in this report, except where we have noted that some children were excluded because of incomplete data for analyses that spanned several timepoints or examined several instruments simultaneously. In particular, children classified by local educational agencies as handicapped are included in the main analytic sample. Children initially tested in Spanish are included with children initially tested in English for some analyses. This practice differs from that of previous analyses and reports; the rationale for both decisions is discussed below.

The rationale for exclusion of children identified as handicapped from the analytic samples in previous reports (cf. for example Granville et al., 1979, Report IX) was concern that handicaps might unduly impair the children's test performance. For this report analyses were conducted to establish the extent and importance of differences between handicapped and nonhandicapped children in the spring 1979 sample; results are summarized in Table 5. Means for the handicapped children are well within acceptable instrument ranges, and standard deviations are quite comparable for the two groups. Subgroups of handicapped children classified as to type of handicap were also compared, with the same results. There appears to be no reason why any of the children classified as handicapped in the spring 1979 PDC sample should be excluded from analyses on test-performance grounds.

Children judged by teachers and observers or examiners to be Spanish-dominant had, in years prior to 1979, been tested in Spanish rather than English. For this reason, in past reports they have been excluded from the main analytic sample and considered in separate analyses. In spring 1979, however, all children were tested in English. It became appropriate, therefore, to consider whether these children could be included within the analytic sample to the extent possible, or should as in past years be treated separately.

Table 4

Spring 1979 Sample Composition and Distribution by Site

				ETHNICITY					SEX		LAN-GUAGE[a]	
		Number in Full Sample	% Handicapped	% Black	% Hispanic	% American Indian/ Native Alaskan	% White	% Asian/Pacific Islander	% Female	% Male	% English	% Spanish
CALIFORNIA	PDC	21	29	5	95	0	0	0	71	29	81	19
	Comp	25	16	4	32	0	24	0	60	40	72	28
COLORADO	PDC	20	0	15	75	0	10	0	35	65	100	0
	Comp	14	0	0	79	0	21	0	36	64	100	0
CONNECTICUT	PDC	35	12	51	29	3	17	0	51	49	77	23
	Comp	36	17	86	11	0	3	0	44	56	100	0
FLORIDA	PDC	31	7	100	0	0	0	0	32	68	100	0
	Comp	26	9	85	12	0	4	0	35	65	100	0
GEORGIA	PDC	27	4	85	0	0	15	0	56	44	100	0
IOWA	PDC	15	0	33	7	0	60	0	47	53	100	0
	Comp	14	0	21	0	7	72	0	50	50	100	0
MARYLAND	PDC	27	22	44	7	0	48	0	37	63	100	0
	Comp	22	27	32	27	0	36	5	50	50	100	0
MICHIGAN	PDC	21	0	57	5	0	38	0	52	48	100	0
	Comp	32	0	75	3	0	22	0	47	53	100	0
TEXAS	PDC	37	17	3	78	0	19	0	38	62	32	68
	Comp	40	25	0	90	0	10	0	53	47	35	65
UTAH	PDC	23	26	0	26	0	74	0	39	61	100	0
	Comp	34	18	3	12	6	79	0	62	38	100	0
WASHINGTON	PDC	25	24	24	0	20	40	16	56	44	100	0
	Comp	26	0	42	0	4	54	0	73	27	100	0
TOTALS BY GROUP	PDC	282	14	40	30	2	27	1	46	54	87	13
	Comp	269	13	37	31	2	30	0	52	48	88	12
TOTALS, ALL GROUPS COMBINED		551	13	39	30	2	29	1	49	51	87	13

[a]Language in which child was tested in the study cohort's Head Start year (1976-1977).

Table 5

Comparison of Handicapped (n=72) and Non-Handicapped (n=466) Children in PDC Spring 1979 Analytic Sample[1]

	Handicapped	Non-handicapped	Difference
% male	67%	49%	--
Age at entry (mo.)	53.5(4.2)	53.8(4.2)	--
Mother's Educ. (yrs.)	10.4(2.8)	10.6(2.4)	--
BSM-English	12.4(3.0)	12.6(3.0)	--
Verbal Fluency	16.1(5.8)	16.4(6.1)	--
POCL-1	38.0(10.3)	42.2(10.3)	(-.4 sd)
POCL-2	14.9(4.7)	15.0(4.6)	--
PIAT-Math	16.5(5.3)	19.1(6.1)	(-.4 sd)
PIAT-Reading	20.9(6.5)	23.1(5.6)	(-.4 sd)

[1] Information on handicap status is not available for 13 children in the spring 1979 analytic sample. Except for sex, table entries are means (standard deviations). Differences are presented as approximate standard deviation unit equivalents, where appropriate and greater than .1 standard deviation.

Comparisons of child outcome scores for children initially tested in Spanish and those initially tested in English, using spring 1979 data, were carried out at site and aggregate levels. The comparisons showed that there was no reason why, in terms of major test-performance criteria, the two groups could not be analyzed together. Table 6 presents site-level data for children initially tested in Spanish and English for selected child outcome variables measured at grade 1 (spring 1979).

Although it seems both desirable and justifiable that a single analytic sample be formed regardless of initial testing language, concern about initial comparability in the longitudinal design of the study makes certain compromises in analysis necessary. Measures prior to 1979 differ for these two groups by language of administration; although the measures are analogous, they are not identical and might be tapping somewhat different psychological constructs. Consequently, repeated measures analyses incorporating data from earlier years, as well as analyses of spring 1979 data using entry-level data as covariates, are performed separately for subgroups initially tested in Spanish or English (cf. Chapter V). However, analyses of sample and instrument characteristics, appearing in this chapter, and outcome analyses restricted to spring 1979 data in the next, are presented for the combined sample of children, including together those initially tested in Spanish and those initially tested in English.

Child Sample Attrition Patterns and Their Effects

Since entry into the Head Start program in fall 1976, the available sample of children for the study has gradually grown smaller. The present section examines patterns of attrition and their effects on the representativeness of the sample of children remaining and on the comparability of PDC and comparison groups remaining in the study at the end of first grade (spring 1979).

Attrition patterns. Table 7 displays the pattern of attrition by site and treatment group from fall 1976 through spring 1979. Overall, the total PDC sample has been reduced by 51%, and the total comparison sample by 52% in the evaluation's first phase, it had been projected that attrition rates for this period would be 46% for the PDC sample and 51% for the comparison sample.

Attrition's effect on sample representativeness. Beyond its effect on sample size, attrition can lead to samples that are not representative of original program groups. For example, if families of higher socioeconomic status are more mobile than families of lower status, the mean socioeconomic status of families remaining in a study such as the present one will decrease over time, perhaps to a point such that the longitudinal sample can no longer be considered to represent the original study sample. Therefore, analyses were conducted to assess changes in representativeness in the sample available to the study as of grade 1 (spring 1979).

Table 6

Means and Standard Deviations for Children Initially Tested in Spanish and English Overall and at Three Sites on Six Child Outcome Variables in the Spring 1979 Analytic Sample[1]

Outcome Variable	California	Connecticut	Texas	Overall
N (Tested in English/ Tested in Spanish)	35/11	68/8	26/51	124/70
BSM-English:				
Tested in English	13.7(2.1)	12.1(2.4)	15.0(1.4)	13.1(2.4)
Tested in Spanish	13.2(2.3)	7.9(3.0)	13.2(2.1)	12.6(2.8)
Verbal Fluency:				
Tested in English	16.0(5.4)	20.1(5.5)	17.4(4.9)	18.4(5.6)
Tested in Spanish	14.8(6.5)	13.6(5.0)	13.8(4.3)	14.0(4.7)
POCL-1:				
Tested in English	36.9(7.4)	42.7(11.4)	44.9(9.5)	41.6(10.4)
Tested in Spanish	44.9(10.4)	37.8(7.1)	41.3(12.5)	41.3(11.8)
POCL-2:				
Tested in English	13.6(3.1)	14.3(5.1)	16.3(4.9)	14.5(4.6)
Tested in Spanish	15.0(4.0)	13.7(5.4)	14.6(5.5)	14.5(5.2)
PIAT-Math:				
Tested in English	17.0(4.0)	17.4(4.8)	23.0(6.6)	18.4(5.5)
Tested in Spanish	16.7(6.1)	14.4(3.3)	17.1(3.8)	16.8 (4.2)
PIAT-Reading				
Tested in English	21.1(6.0)	22.1(6.4)	26.2(5.1)	22.7(6.3)
Tested in Spanish	19.8(4.2)	15.3(7.0)	22.8(4.4)	21.6(5.1)

[1]Numbers of cases may vary slightly within sites from one measure to another. Table entries are presented as: mean (standard deviation).

Table 7

Summary of Year-to-Year Attrition[1]

	PDC					COMPARISON				
	Fall 1976 N	Spring 1977 N	Spring 1978 N	Spring 1979 N	1976-1979 % Drop	Fall 1976 N	Spring 1977 N	Spring 1978 N	Spring 1979 N	1976-1979 % Drop
rnia	44	37	25	21	53	40	32	27	25	38
do	55	50	25	20	64	32	25	22	14	56
ticut	51	47	32	35	31	62	60	46	36	42
a	47	39	40	31	34	39	35	34	26	33
a	46	43	30	27	41	----------No Comparison Sample----------				
	50	43	25	15	70	54	48	22	14	74
nd	44	41	32	27	39	58	47	30	22	62
an	66	58	31	21	68	64	58	51	32	50
	64	54	46	37	42	57	55	42	40	30
	48	45	28	23	52	81	73	60	34	58
gton	58	52	35	25	57	76	61	47	26	66
	573	509	349	282	51	563	494	381	269	52

on group designation at KDG.

Table 8 presents entry-level values on background characteristics and test scores for children in the original sample, children remaining in the sample as of spring 1979, and children lost to the study over the period 1976-1979. Test data for children tested at entry in English and in Spanish are presented separately, given the difficulty of interpreting combined raw means for these two groups taken together.

Table 9 summarizes the results of comparisons between remaining and departed child subsamples for the children initially tested in English, for those initially tested in Spanish and for the full analytic sample. In order to be able to compare remaining and departed subsamples for the full analytic sample, corresponding test scores for each group tested at entry in a different language were separately standardized and polled.

As Tables 8 and 9 show, significant differences between the sample remaining and the sample departed over the study's three years appear both for background characteristics and for test scores. The ethnic composition of the study sample has changed significantly from entry to spring 1979: the proportion of children of black and Hispanic origin has increased, while the proportion of children of white origin has decreased. The proportion of children with prior preschool experience has increased. For the subsample of children initially tested in English, changes in entry-level test scores appear as significant differences between the sample remaining and the sample departed for the WPPSI, for which the sample remaining has slightly lower mean levels, and for the test of Arm Coordination, for which the sample remaining has slightly higher mean levels. The subsample of children tested initially in Spanish shows significant differences for three tests: the BSM-Spanish, Arm Coordination and Draw-A-Child. In all three cases, the sample remaining has higher scores than the sample departed. For the full analytic sample, significant differences between children departed and those remaining appear for two measures: Verbal Memory-1 and Arm Coordination. In both cases, the sample remaining is higher in mean levels than the sample departed.

Entry-level differences in test scores were also examined in a multivariate analysis of variance considering all variables simultaneously for each sample of children. Separate multivariate analyses were performed for children initially tested in English, children tested in Spanish, and the full analytic sample. All ten entry-level test scores were included as dependent variables. Sample size for these comparisons ranged between 92% and 96% of the total available, since only children with data on all ten variables could be incorporated in these analyses. Multivariate tests revealed statistically significant differences between remaining and departed samples for the subsample initially tested in English ($F = 2.03$, 10 and 966 degrees of freedom; $p = .028$) and for the total analytic sample ($F = 1.86$, 10 and 1,073 degrees of freedom; $p = .047$), but not for the subsample initially tested in Spanish ($F = 1.14$, 10 and 96 degrees of freedom; $p = .33$).

Table 8

Representativeness of Remaining Study Sample:
Comparison of Sample Remaining and Sample Departed as of
Spring 1979 on Background Characteristics and Entry Level Test Scores

Full Analytic Sample			
Background Characteristics	Original Sample[1]	Spring 1979: Sample Remaining	Spring 1979: Sample Departed
N (maximum)	1136	551	585
Ethnicity(%)* Black Hispanic American Indian/ Native American White Asian/Pacific Islander	 36% 27 2 33 2	 39% 30 2 28 1	 34% 24 2 37 3
Sex(%) Male Female	 50% 50	 51% 49	 49% 51
Prior Preschool(%)* Yes No	 15% 85	 17% 83	 13% 87
Age (months, at entry) Number of Siblings Mother's Education (years)	53.8 1.91 10.7	53.8 1.97 10.6	53.8 1.85 10.8

[1]Children from the West Virginia site, which dropped out of the longitudinal study in summer 1978, are not included.

*Difference on this variable between remaining and departed groups significant with $p<.10$.

Note: Table entries are group means, except as indicated.

Table 8
(continued)

Sample Initially Tested in English			
Fall 1976 Test Scores	Original Sample[1]	Spring 1979: Sample Remaining	Spring 1979: Sample Departed
N(approximate)	1019	481	538
BSM-English	9.11	9.00	9.21
WPPSI*	4.94	4.72	5.13
Verbal Fluency	5.97	6.12	5.83
Verbal Memory-1	13.10	13.40	12.84
Verbal Memory-3	2.64	2.60	2.67
Arm Coordination*	3.39	3.60	3.21
Draw-A-Child	3.85	3.73	3.97
PIPS	1.99	1.92	2.06
POCL-1: "Task Orientation"	32.88	33.22	32.57
POCL-2: "Sociability"	12.84	12.93	12.75

[1]Children from the West Virginia site, which dropped out of the longitudinal study in summer 1978, are not included.

*Difference on this variable between remaining and departed groups significant with $p<.10$.

Note: Table entries are group means.

Table 8
(continued)

Sample Initially Tested in Spanish			
Fall 1976 Test Scores	Original Sample[1]	Spring 1979: Sample Remaining	Spring 1979: Sample Departed
N(approximate)	117	70	47
BSM-Spanish*	10.91	11.71	9.73
WPPSI	6.71	7.17	6.02
Verbal Fluency	4.06	4.39	3.56
Verbal Memory-1	12.07	12.89	10.85
Verbal Memory-3	2.02	2.22	1.72
Arm Coordination*	3.90	4.38	3.13
Draw-A-Child*	4.10	4.53	3.44
PIPS	1.70	1.86	1.44
POCL-1: "Task Orientation"	29.64	30.10	28.96
POCL-2: "Sociability"	11.16	11.19	11.13

[1]Children from the West Virginia site, which dropped out of the longitudinal study in summer 1978, are not included.

*Difference on this variable between remaining and departed groups significant with $p<.10$.

Note: Table entries are group means.

Table 9

Representativeness of Remaining Study Sample:
Univariate and Multivariate Test Score Comparisons of
Samples Remaining and Departed, by Language of
Initial Test and Grouped Together (Full Analytic Sample)

Fall 1976 Test Score	Language of Initial Test:		Full Analytic Sample
	English	Spanish	
Univariate Comparison *N (maximum)*	1019	117	1136
BSM-English	n.d.[1]	----	----
BSM-Spanish	----	.05(R>D)	----
BSM-both tests	----	----	n.d.
WPPSI	.06(D>R)	n.d.	n.d.
Verbal Fluency	n.d.	n.d.	n.d.
Verbal Memory-1	n.d.	n.d.	.09(R>D)
Verbal Memory-3	n.d.	n.d.	n.d.
Arm Coordination	.04(R>D)	.07(R>D)	.02(R>D)
Draw-A-Child	n.d.	.04(R>D)	n.d.
PIPS	n.d.	n.d.	n.d.
POCL-1: "Task Orientation"	n.d.	n.d.	n.d.
POCL-2: "Sociability"	n.d.	n.d.	n.d.
Multivariate Comparison *N*	977 (95.9%)	107 (91.5%)	1084 (95.4%)
All Tests Together: Significance	.03	n.d.	.05

[1]Table entries are: n.d. if there is no significant difference between samples remaining and departed; or significance level of difference and direction as x(R>D) where R: sample remaining, and D: sample departed.

In sum, attrition of children from the sample between fall 1976 and spring 1979 has altered its composition to some degree. This finding differs from the results of similar analyses conducted on samples remaining at the end of the Head Start year (spring 1977, reported in Interim Report VII, Volume 3, pp. 17-22) and on samples remaining at the end of the children's kindergarten year (spring 1978, reported in Interim Report IX, pp. 16-21).

The finding does not appear to have any major implications, however. While both multivariate and univariate analyses indicate that the sample remaining at the end of grade 1 is not quite the same as the original sample for the study, differences are neither consistent nor substantial and are not expected to alter the validity of study findings. For one thing, there is no clearly discernible trend in the direction of sample changes. Changes in background characteristics cannot be readily interpreted as changes in socioeconomic status: minority ethnic group composition has increased, but mother's education has remained unchanged and the proportion of children with preschool experience prior to Head Start has increased. Changes in test score entry levels also fail to show any consistent trend: of the ten differences between mean levels for each of the subsamples initially tested in English and the full analytic sample, half favor the sample remaining and half favor the sample departed. Further, differences between groups are small (about .1 standard deviation units or less in magnitude of differences between group means).

Attrition's effects on group comparability. When PDC and comparison children entered Head Start, they were found to have very nearly the same background characteristics and test scores (Interim Report VI, pp. 30-36). This is to say, PDC and comparison samples differed only in the treatment groups to which they were assigned. By the end of the kindergarten year (spring 1978), the samples of PDC and comparison children remaining in the study were still quite similar with respect to background and entry-level test scores (Interim Report IX, pp. 16-21). Here we present a series of univariate and multivariate analyses of background characteristics and entry-level test scores comparing PDC and comparison groups at the end of first grade (spring 1979).

Table 10 compares the background characteristics and entry-level test scores of PDC and comparison children remaining in the sample in spring 1979. As in the preceding section, test scores are reported separately for children initially tested in Spanish and in English.

Table 11 summarizes statistical tests on the comparisons presented in Table 10, together with the results of analyses for the full analytic sample, combining children initially tested in Spanish and English. The latter analysis was accomplished by separately standardizing the test scores for children tested in Spanish and English, then pooling the standardized scores for analysis.

Table 10

Comparability of Remaining PDC and Comparison Children on Baseline (Fall 1976) Characteristics

Full Analytic Sample		
Background Characteristics	PDC Sample Spring 1979	Comparison Sample Spring 1979
N (maximum)	*280*	*271*
Ethnicity (%)		
Black	41%	36%
Hispanic	29	32
American Indian/Native American	2	2
White	27	30
Asian/Pacific Islander	1	0
Sex (%)		
Male	54%	49%
Female	46	52
Prior Preschool (%)		
Yes	16%	19%
No	84	81
Age (months)	53.68	53.87
Number of Siblings	1.97	1.95
Mother's Education	10.70	10.49

Note: Table entries are group means, except as indicated.

Table 10
(continued)

Comparability of Remaining PDC and Comparison Children
on Baseline (Fall 1976) Characteristics:
Sample Initially Tested in English

Sample Initially Tested in English		
Fall 1976 Test Scores	PDC Sample Spring 1979	Comparison Sample Spring 1979
N (maximum)	*243*	*237*
BSM-English	8.72	9.32
WPPSI *	4.36	5.10
Verbal Fluency	6.16	6.10
Verbal Memory-1	13.11	13.70
Verbal Memory-3	2.56	2.68
Arm Coordination	3.39	3.83
Draw-A-Child	3.58	3.90
PIPS	1.90	1.97
POCL-1: "Task Orientation"	32.70	33.61
POCL-2: "Sociability" *	12.42	13.46

*PDC-comparison group difference on this variable significant with $p<.10$ (two-tailed).

Note: Table entries are group means.

Table 10
(continued)

Comparability of Remaining PDC and Comparison Children
on Baseline (Fall 1976) Characteristics:
Sample Initially Tested in Spanish

Sample Initially Tested in Spanish		
Fall 1976 Test Scores	PDC Sample Spring 1979	Comparison Sample Spring 1979
N (maximum)	*37*	*34*
BSM-Spanish	11.68	11.46
WPPSI	7.43	6.85
Verbal Fluency	3.70	5.00
Verbal Memory-1	12.30	13.24
Verbal Memory-3	2.28	2.09
Arm Coordination	4.37	4.30
Draw-A-Child	4.35	4.59
PIPS	1.78	1.88
POCL-1: "Task Orientation"	30.81	29.65
POCL-2: "Sociability"	11.16	11.24

Note: Table entries are group means.

Table 11

Comparability of Remaining PDC and Comparison Children: Univariate and Multivariate Test Score Comparisons of PDC and Comparison Groups, by Language of Initial Test and Grouped Together (Full Analytic Sample)[1]

Fall 1976 Test Scores	Language of Initial Test:		Full Analytic Sample
	English	Spanish	
Univariate Comparison *N (maximum)*	*481*	*70*	*551*
BSM-English	n.d.[1]	----	----
BSM-Spanish	----	n.d.	----
BSM-both tests	----	----	n.d.
WPPSI	.02(C>P)	n.d.	.05(C>P)
Verbal Fluency	n.d.	n.d.	n.d.
Verbal Memory-1	n.d.	n.d.	n.d.
Verbal Memory-3	n.d.	n.d.	n.d.
Arm Coordination	n.d.	n.d.	n.d.
Draw-A-Child	n.d.	n.d.	n.d.
PIPS	n.d.	n.d.	n.d.
POCL-1: "Task Orientation"	n.d.	n.d.	n.d.
POCL-2: "Sociability"	.003(C>P)	n.d.	.006(C>P)
Multivariate Comparison *N*	463 (96.3%)	65 (92.9%)	528 (95.8%)
All Tests Together: Significance	.019	n.d.	.028

[1]Table entries are: n.d. if there is no significant difference between PDC and comparison group children; or significance level of difference and direction as $\alpha(P>C)$ where P: PDC group and C: comparison group.

As the tables show, there are no significant differences between groups on background characteristics. There are, however, differences between groups on test scores. For the sample of children initially tested in English, significant differences in entry-level test scores appear for the WPPSI and POCL-2. What is more, differences between the PDC and comparison groups show a consistent direction: for nine of ten measures, and both of the statistically significant ones, absolute differences in entry-level scores favor the comparison group. A multivariate analysis of variance applied to this sample confirms this finding, showing a significant difference between PDC and comparison groups at entry ($F = 2.16$, d.f.: 10, 452; $p = .020$). No such difference appears for the subsample of children initially tested in Spanish. When both subsamples are combined (full analytic sample) to test for a difference in initial levels between treatment groups, the findings replicate those for the subsample initially tested in English. There are multivariate differences between the two groups on entry-level test scores, with all differences consistently favoring the comparison group. Two tests show significant differences on univariate comparison: the WPPSI and the POCL-2. Sample sizes for the multivariate comparisons range between 92 and 96% of the total available samples for univariate comparisons.

In sum, by spring 1979, attrition had affected the initial comparability of the treatment groups. This finding differs from findings at entry and spring 1978. Although there has been no significant change in the background characteristics of the two treatment groups, they are no longer strictly comparable on entry-level test scores. Further, there appears to be a definite directional tendency to the difference between groups, with the comparison group higher on entry-level test scores than the PDC group. Differences between treatment group mean levels are not large (amounting to one-quarter of a standard deviation or less), but they are consistent.

The implications of entry-level differences between groups for child outcome analyses are not trivial. Covariance adjustment for entry-level differences becomes necessary for the group initially tested in English, while such adjustment appears unnecessary for children initially tested in Spanish. The issue of adjustment for entry-level differences between treatment groups is dealt with in Chapter V, where the analyses of child outcome data for the spring 1979 analytic sample are discussed.

Characteristics of the Child Measures

Table 12 provides a key to the abbreviated names of child outcome measures used in this report. Descriptions of the measures and of the procedures by which summary scores are obtained can be found in Appendix A.

Table 12

Key to Abbreviated Names of Child Outcome Measures[1]

BSM-English	English language version of the Bilingual Syntax Measure
BSM-Spanish	Spanish language version of the Bilingual Syntax Measure
Verbal Fluency	Verbal Fluency subtest of the McCarthy Scales of Children's Abilities
PIAT-Math	Mathematics subtest of the Peabody Individual Achievement Test
PIAT-Reading	Reading subtest of the Peabody Individual Achievement Test
PIPS	Preschool Interpersonal Problem Solving Test
POCL-1	Task Orientation subscale of the Pupil Observation Checklist
POCL-2	Sociability subscale of the Pupil Observation Checklist
CRS-1	Self-Assurance subscale of the Child Rating Scale
CRS-2	Aggressiveness subscale of the Child Rating Scale
CRS-3	Dependence subscale of the Child Rating Scale
CRS-4	Academic Motivation subscale of the Child Rating Scale
CI-1	Attitude Toward Teacher subscale (Part 1) of the Child Interview
CI-2	Interest in Reading subscale (Part 2) of the Child Interview

[1]Child measures and summary scores are described in detail in Appendix A.

Score Distributions

Table 13 presents ranges, means, medians and standard deviations of the summary scores for each child outcome measure included in the grade 1 (spring 1979) battery. These statistics are computed for data based on all PDC and comparison children in the analytic sample. Score distributions for each summary measure are graphically displayed in histograms appearing as Figures 3-15.

Score Stability

As would be expected of measures sensitive to children's learning and development, mean levels for the summary scores obtained on more than one occasion between Head Start and first grade show increases over time where appropriate. Table 14 presents the mean levels and sample sizes available at all timepoints for which the same instruments are administered as those used in grade 1 (spring 1979). The data in this table exclude those children tested in English in spring 1979 who had been tested in Spanish on one or more occasions in previous years, since they in effect were not tested with the same instruments at all times.

Another perspective on score stability is provided by summary score correlations between one testpoint and another, indicating the degree to which the relative positions of children's scores change over time. Table 15 presents, again for those children for whom the same instruments were administered over time, the occasion-to-occasion correlations for each measure in the grade 1 (spring 1979) battery administered more than once.

Internal Consistency

Table 16 reports coefficients of internal consistency (Cronbach's alpha) for the summary scores of each measure in the spring 1979 battery. All but one of these coefficients exceed the criterion of .65 established as the cut-off for acceptable item homogeneity in previous reports. The single exception is the "attitude toward teacher" scale from part 1 of the Child Interview, for which a coefficient of .54 was obtained for the study sample at spring first grade.

Since the Child Interview was not administered prior to spring 1979, no comparison with coefficients from previous PDC data collections is possible. When a slightly different version of the instrument was administered to similar first grade children in a longitudinal study of Head Start impacts (see Appendix A, Attitude Toward Teacher), similarly low internal consistency was found. However, in the absence of test-retest findings indicating short-term instability and in the absence of clearly

Table 13

Descriptive Statistics for Spring Grade 1 (1979)
Child Outcome Measure Summary Scores

Measure	N	Range	Mean	Median	Standard Deviation
BSM-English	546	1-18	12.54	13.78	3.03
Verbal Fluency	546	2-44	16.24	15.93	6.07
PIAT-Math	535	6-41	18.70	19.34	6.04
PIAT-Reading	516	5-46	22.82	23.63	5.74
PIPS	545	0-8	4.17	4.74	1.65
POCL-1 Task Orientation	513	11-56	41.59	42.03	10.41
POCL-2 Sociability	513	3-21	14.96	15.13	4.62
CRS-1 Self-Assurance	510	11-55	35.18	34.75	7.52
CRS-2 Aggressiveness	521	4-20	10.06	10.40	3.62
CRS-3 Dependence	522	2-10	5.47	6.20	1.93
CRS-4 Academic Motivation	517	3-15	9.75	9.78	3.18
CI-1 Attitude Toward Teacher	544	1-5	2.30	2.35	0.68
CI-2 Interest in Reading	529	3-15	9.42	9.85	3.20

Figure 3

Distribution of Spring Grade 1 (1979) BSM-English Scores

N=546

HISTOGRAM

MIDPOINT	COUNT	(EACH X= 5)
1.0000	1	+X
3.0000	3	+X
5.0000	6	+XX
7.0000	29	+XXXXXX
9.0000	63	+XXXXXXXXXXXX
11.000	73	+XXXXXXXXXXXXXX
13.000	111	+XXXXXXXXXXXXXXXXXXXXX
15.000	173	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
17.000	85	+XXXXXXXXXXXXXXX
19.000	2	+X

(INTERVAL WIDTH= 2.0000)

Figure 4

Distribution of Spring Grade 1 (1979) Verbal Fluency Scores

N=546

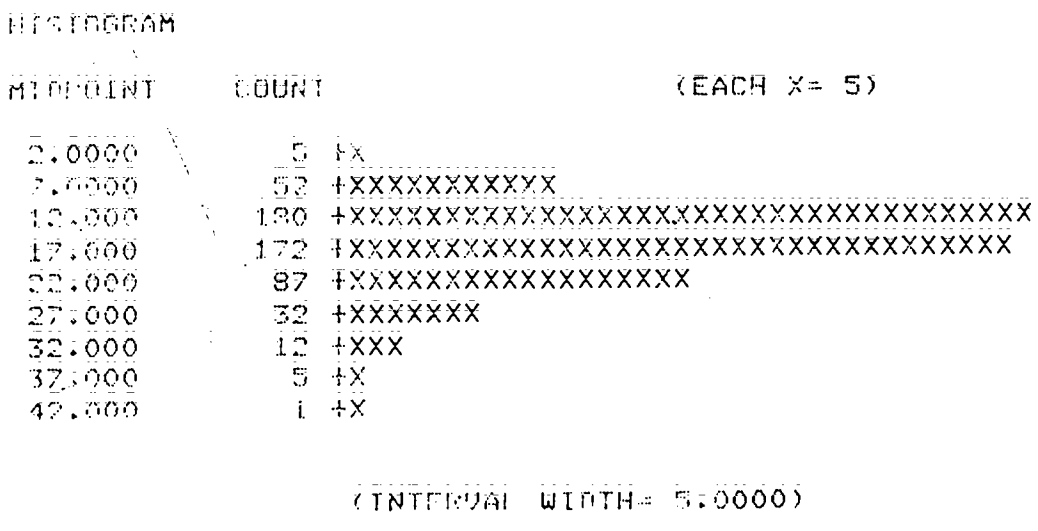


Figure 5

Distribution of Spring Grade 1 (1979) Summary Scores
for PIAT-Math Scores

N=535

HISTOGRAM

MIDPOINT	COUNT	(EACH X= 5)
6.0000	2	+X
11.000	101	+XXXXXXXXXXXXXXXXXXXX
16.000	195	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
21.000	144	+XXXXXXXXXXXXXXXXXXXXXXXXXXXX
26.000	57	+XXXXXXXXXXXX
31.000	19	+XXXX
36.000	13	+XXX
41.000	4	+X

(INTERVAL WIDTH= 5.0000)

Figure 6

Distribution of Spring Grade 1 (1979) Summary Scores
for PIAT-Reading Scores

N=516

HISTOGRAM

MIDPOINT	COUNT	(EACH X= 8)
5.0000	3	+X
10.000	13	+XXX
15.000	42	+XXXXXXX
20.000	218	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
25.000	152	+XXXXXXXXXXXXXXXXXXXXXXXXX
30.000	59	+XXXXXXXXXX
35.000	18	+XXX
40.000	7	+XX
45.000	4	+X

(INTERVAL WIDTH= 5.0000)

Figure 7

Distribution of Spring Grade 1 (1979) Summary Scores
for PIPS Scores

N=545

HISTOGRAM

MIDPOINT	COUNT	(EACH X= 3)
0.	3	+X
1.0000	34	+XXXXXXXXXXX
2.0000	51	+XXXXXXXXXXXXXXXX
3.0000	101	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
4.0000	113	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
5.0000	118	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
6.0000	90	+XXXXXXXXXXXXXXXXXXXXXXXXXXXX
7.0000	29	+XXXXXXXXXX
8.0000	6	+XX

(INTERVAL WIDTH= 1.0000)

Figure 8

Distribution of Spring Grade 1 (1979) Summary Scores
for POCL-1 Scores

N=513

HISTOGRAM

MIDPOINT	COUNT	(EACH X= 3)
11.000	2	+X
16.000	8	+XXX
21.000	9	+XXX
26.000	36	+XXXXXXXXXXX
31.000	77	+XXXXXXXXXXXXXXXXXXXXXXXXX
36.000	70	+XXXXXXXXXXXXXXXXXXXXXX
41.000	72	+XXXXXXXXXXXXXXXXXXXXXX
46.000	78	+XXXXXXXXXXXXXXXXXXXXXXXX
51.000	74	+XXXXXXXXXXXXXXXXXXXXXXX
54.000	87	+XXXXXXXXXXXXXXXXXXXXXXXXXXX

(INTERVAL WIDTH= 5.0000)

Figure 9

Distribution of Spring Grade 1 (1979) Summary Scores
for POCL-2 Scores

N=513

HISTOGRAM

MIDPOINT	COUNT	(EACH X= 3)
3.0000	4	+XX
6.0000	32	+XXXXXXXXXX
9.0000	55	+XXXXXXXXXXXXXXXXXX
12.000	111	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
15.000	99	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
18.000	105	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
21.000	107	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

(INTERVAL WIDTH= 3.0000)

Figure 10

Distribution of Spring Grade 1 (1979) Summary Scores
for CRS-1 Scores

N=510

```
HISTOGRAM

MIDPOINT     COUNT                          (EACH X= 4)

 11.000          1 +X
 16.000          5 +XX
 21.000         19 +XXXXX
 26.000         60 +XXXXXXXXXXXXXXX
 31.000        149 +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 36.000        136 +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 41.000         64 +XXXXXXXXXXXXXXXX
 46.000         45 +XXXXXXXXXXXX
 51.000         23 +XXXXXX
 56.000          8 +XX

                    (INTERVAL WIDTH= 5.0000)
```

63

Figure 11

Distribution of Spring Grade 1 (1979) Summary Scores
for CRS-2

N=521

HISTOGRAM

MIDPOINT	COUNT	(EACH X= 3)
4.0000	35	+XXXXXXXXXXX
6.0000	63	+XXXXXXXXXXXXXXXXXXXX
8.0000	91	+XXXXXXXXXXXXXXXXXXXXXXXXXXXX
10.000	102	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
12.000	103	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
14.000	64	+XXXXXXXXXXXXXXXXXXXXX
16.000	39	+XXXXXXXXXXXXX
18.000	17	+XXXXXX
20.000	7	+XXX

(INTERVAL WIDTH= 2.0000)

Figure 12

Distribution of Spring Grade 1 (1979) Summary Scores
for CRS-3

N=522

HISTOGRAM

MIDPOINT	COUNT	(EACH X= 4)
2.0000	54	+XXXXXXXXXXXXX
3.0000	36	+XXXXXXXXX
4.0000	67	+XXXXXXXXXXXXXXX
5.0000	75	+XXXXXXXXXXXXXXXXX
6.0000	145	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
7.0000	69	+XXXXXXXXXXXXXXXXX
8.0000	50	+XXXXXXXXXXXX
9.0000	17	+XXXXX
10.000	9	+XXX

(INTERVAL WIDTH= 1.0000)

Figure 13

Disbribution of Spring Grade 1 (1979) Summary Scores
for CRS-4
N=517

HISTOGRAM

MIDPOINT	COUNT	(EACH X= 5)
3.0000	17	+XXXX
5.0000	32	+XXXXXXX
7.0000	65	+XXXXXXXXXXXXX
9.0000	163	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
11.000	79	+XXXXXXXXXXXXXXX
13.000	84	+XXXXXXXXXXXXXXXX
15.000	77	+XXXXXXXXXXXXXXX

(INTERVAL WIDTH= 2.0000)

Figure 14

Distribution of Spring Grade 1 (1979)
Child Interview Part 1 Scores

N=544

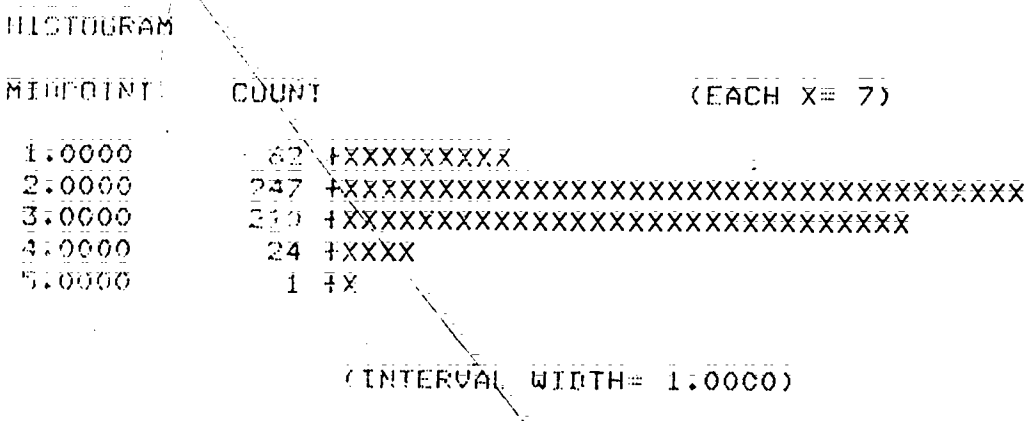


Figure 15

Distribution of Spring Grade 1 (1979)
Child Interview Part 2 Scores

N=529

HISTOGRAM

MIDPOINT	COUNT	(EACH X= 4)
3.0000	36	+XXXXXXXXX
5.0000	27	+XXXXXXX
7.0000	79	+XXXXXXXXXXXXXXXXXXX
9.0000	132	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
11.000	87	+XXXXXXXXXXXXXXXXXXXX
13.000	119	+XXXXXXXXXXXXXXXXXXXXXXXXXXXX
15.000	49	+XXXXXXXXXXXX

(INTERVAL WIDTH= 2.0000)

Table 14

Means of Child Outcome Measures Repeated Over Time and
Administered in Spring 1979 for all Children[1] Still Present in Spring 1979

Fall 1976			Spring 1977			Spring 1978			Spring 1979		
	N	Mean		N	Mean		N	Mean		N	Mean
BSM-English	480	9.00	BSM-English	463	9.73	BSM-English	474	11.62	BSM-English	481	12.54
BSM-Spanish	67	11.57	BSM-Spanish	69	11.72	BSM-Spanish	39	12.87	BSM-Spanish	70	14.63
Verbal Fluency	473	6.13	Verbal Fluency	462	8.81	Verbal Fluency	445	14.31	Verbal Fluency	481	16.56
PIPS	474	1.93	PIPS	462	2.55	PIPS	445	3.68	PIPS	480	4.24
POCL-1	481	33.15	POCL-1	433	33.58	POCL-1	414	37.02	POCL-1	455	41.63
POCL-2	481	12.93	POCL-2	433	13.25	POCL-2	414	13.29	POCL-2	455	15.01
			CRS-1	399	35.22	CRS-1	398	34.85	CRS-1	450	35.20
			CRS-2	408	11.62	CRS-2	423	11.50	CRS-2	459	10.13
			CRS-3	412	5.80	CRS-3	421	5.51	CRS-3	457	5.42
									CRS-4	453	9.76
						PIAT-Math	429	13.23	PIAT-Math	472	18.99
						PIAT-Reading	416	15.25	PIAT-Reading	457	23.00
									CI-1	479	2.31
									CI-2	461	9.73

KEY:

POCL-1: "Task Orientation"
POCL-2: "Sociability"
CRS-1: "Self-Assurance"
CRS-2: "Aggressiveness"
CRS-3: "Dependence"
CRS-4: "Academic Motivation
CI-1: "Attitude Toward Teacher"
CI-2: "Interest in Reading"

[1] Approximately 70 children tested in spring 1979 in English but who had been tested in Spanish in earlier years are excluded from this table.

Table 15

Stability (Test-Retest Correlations) of the Child Measures at Four Timepoints: Fall 1976, Spring 1977, Spring 1978, and Spring 1979

Measures	*N*	F76-S77 r_{tt}	F76-S78 r_{tt}	F76-S79 r_{tt}	S77-S78 r_{tt}	S77-S79 r_{tt}	S78-S79 r_{tt}
COGNITIVE-LANGUAGE							
BSM-English	431	.63	.45	.46	.56	.56	.48
BSM-Spanish	35	.61	.44	.58	.74	.74	.65
Verbal Fluency	424	.48	.36	.25	.54	.36	.44
PIAT-Math[a]	424						.60
PIAT-Reading[a]	398						.56
SOCIAL-EMOTIONAL							
PIPS	432	.37	.18	.23	.30	.31	.29
POCL-1: "Task Orientation"	377	.39	.15	.28	.35	.35	.29
POCL-2: "Sociability"	377	.41	.19	.29	.36	.35	.39
CRS-1: "Self-Assurance"[b]	344				.38	.37	.47
CRS-2: "Aggressiveness"[b]	495				.31	.31	.46
CRS-3: "Dependence"[b]	497				.05	-.03	.14

[a]The PIAT was first administered in spring 1978.

[b]The Child Rating Scale was not administered in fall 1976; thus, test-retest coefficients can be computed only for spring timepoints.

Table 16

Reliability of Child Measures[a]: Cronbach's Alpha
(Internal Consistency) PDC Spring 1979 Data

Measures	Full Analytic Sample	
COGNITIVE-LANGUAGE	*N*	α
BSM-English	542	.75
BSM-Spanish[b]	90	.82
Verbal Fluency	551	.76
LEARNING ATTITUDES		
Child Interview - Part I	544	.54
Child Interview - Part II	529	.94
SOCIAL-EMOTIONAL		
POCL-1: "Task Orientation"	517	.94
POCL-2: "Sociability"	517	.89
CRS-1: "Self-Assurance"	516	.78
CRS-2: "Aggressiveness"	527	.76
CRS-3: "Dependence"	528	.77
CRS-4: "Academic Motivation"	523	.91

[a]Three instruments are not included: the Preschool Interpersonal Problem Solving Test and the Peabody Individual Achievement Test (Reading and Math subtests) do not lend themselves to computation of alpha.

[b]Texas, California and Connecticut only.

more effective measures of young children's attitudes toward school and teachers, it seems unwise to reject part 1 of the Child Interview out of hand. The component scales have been widely used and have been judged to have at least face validity by many researchers and educators. Unless new information comes to light clearly indicating the unsuitability of the measure for the purposes of this research, it is recommended that the "attitude toward teacher" scale be retained and the scores analyzed, keeping in mind that there may be more "noise" than "signal" for this particular measure's summary score.

Correlations Among Summary Scores

Further documentation of the psychometric properties of the measures is provided by intercorrelations of all summary scores for the total spring 1979 sample, in Table 17.

Factor Structure of Summary Scores

In Table 18, the relationships in the intercorrelation matrix are reduced to a smaller number of common factors for the spring 1979 child outcome battery. The factor analysis was accomplished by a varimax rotation of the four factors identified through a principal components analysis as having eigenvalues greater than one.

The factor structure is somewhat ambiguous, with several measures loading moderately (loading levels of .3 or greater) on more than one factor. The proportion of variance accounted for by the first three rotated factors is closely similar, ranging between 16 and 19% of the total variance. The two POCL scales (Task Orientation and Sociability) load heavily on the first factor, while the CI-2 (Interest in Reading) and the PIPS evidence more moderate loadings. Three of the four CRS scales (Self-Assurance, Dependence and Academic Motivation) load heavily on the second factor. The BSM-English, PIAT-Math, and PIAT-Reading load strongly, and Verbal Fluency loads moderately, on the third factor, suggesting that this factor might be interpreted as an index of general academic achievement. The CI-1 (Attitude Toward Teacher) and CRS-2 (Aggressiveness) evidence moderate to strong loadings on the fourth factor; the positive loadings of these factors indicate "negative attitude toward teacher" and "greater aggressiveness/competitiveness," respectively. Because several measures load moderately on two or more factors, interpretation of the factor structure is uncertain, particularly if referred back to the intercorrelation matrix in Table 17.

Table 17

Intercorrelations of Spring 1979 Child Measures

	BSM-ENGLISH	VERBAL FLUENCY	PIAT-MATH	PIAT-READING	PTPS	PDCL1	PDCL2	CBS.1	CBS.2	CBS.3	CBS.4	CBS.5	CBS.6
BSM-ENGLISH	1.0000 (546)	.2306 (546)	.3497 (535)	.3875 (516)	.1676 (545)	.1935 (513)	.1151 (513)	.1915 (510)	.0529 (520)	-.0658 (521)	[illegible] (516)	[illegible] (544)	[illegible] (529)
VERBAL FLUENCY	.2306 (546)	1.0000 (546)	.3062 (535)	.2243 (516)	.2593 (545)	.2575 (513)	.1716 (513)	.2140 (510)	.1343 (520)	-.0530 (521)	.1015 (516)	-.0282 (544)	.1923 (529)
PIAT-MATH	.3497 (535)	.3062 (535)	1.0000 (535)	.5743 (516)	.1537 (534)	.3330 (503)	.1583 (503)	.2888 (499)	.0346 (510)	-.1958 (510)	.2786 (505)	-.0502 (533)	.2794 (519)
PIAT-READING	.3875 (516)	.2243 (516)	.5743 (516)	1.0000 (516)	.1031 (516)	.3580 (486)	.1748 (486)	.4008 (482)	-.0559 (493)	-.2628 (494)	.4040 (488)	-.0421 (515)	.3789 (502)
PTPS	.1676 (545)	.2593 (545)	.1537 (534)	.1031 (516)	1.0000 (545)	.3231 (513)	.2873 (513)	.1322 (510)	.0488 (520)	-.0777 (521)	.1608 (516)	.0572 (544)	.2126 (529)
PDCL1	.1935 (513)	.2575 (513)	.3330 (503)	.3580 (486)	.3231 (513)	1.0000 (513)	.6509 (513)	.2562 (482)	.0443 (491)	-.0988 (492)	.2482 (489)	.0011 (513)	.4222 (499)
PDCL2	.1151 (513)	.1716 (513)	.1583 (503)	.1748 (486)	.2873 (513)	.6509 (513)	1.0000 (513)	.0727 (482)	.1306 (491)	-.0213 (492)	.0462 (489)	-.0811 (513)	.2406 (499)
CBS.1	.1915 (510)	.2140 (510)	.2888 (499)	.4008 (482)	.1322 (510)	.2562 (482)	.0727 (482)	1.0000 (510)	-.1704 (505)	-.4394 (502)	.8360 (500)	.0095 (509)	.2624 (495)
CBS.2	.0529 (520)	.1343 (520)	.0346 (510)	-.0559 (493)	.0488 (520)	.0443 (491)	.1306 (491)	-.1704 (505)	1.0000 (521)	.3061 (513)	-.0917 (508)	.0221 (519)	.0159 (505)
CBS.3	[illegible] (521)	-.0530 (521)	-.1958 (510)	-.2628 (494)	-.0777 (521)	-.0988 (492)	-.0213 (492)	-.4394 (502)	.3061 (513)	1.0000 (522)	-.3764 (510)	-.0882 (520)	.1193 (507)
CBS.4	.1077 (516)	.1815 (516)	.2786 (505)	.4040 (488)	.1608 (516)	.2482 (489)	.0462 (489)	.8360 (500)	-.0917 (508)	-.3764 (510)	1.0000 (517)	.0344 (515)	.2455 (500)
CBS.5	-.0561 (544)	-.0282 (544)	-.0502 (533)	-.0421 (515)	.0572 (544)	.0011 (513)	-.0811 (513)	.0095 (509)	.0221 (519)	-.0882 (520)	.0344 (515)	1.0000 (544)	-.0766 (529)
CBS.6	.1814 (529)	.1923 (529)	.2794 (519)	.3789 (502)	.2126 (529)	.4222 (499)	.2406 (499)	.2624 (495)	-.0159 (505)	-.1193 (505)	.2455 (500)	-.0766 (529)	1.0000 (529)

Table 18

Factor Analysis[1] of Scores on Child Measures
Spring 1979

N = 425	Factor Loading of Child Measures *(highest loading italicized)*			
Child Measure	Factor 1	Factor 2	Factor 3	Factor 4
BSM-English	.03	.05	*.73*	.02
Verbal Fluency	.29	.05	*.41*	.37
PIAT-Math	.15	.25	*.71*	.08
PIAT- Reading	.18	.36	*.69*	-.11
PIPS	*.55*	.16	.03	.35
POCL-1: "Task Orientation"	*.84*	.13	.22	.01
POCL-2: "Sociability"	*.86*	-.08	.04	-.02
CI-2: "Interest in Reading"	*.48*	.22	.34	-.17
CRS-1: "Self-Assurance"	.10	*.83*	.26	-.05
CRS-3: "Dependence"	-.08	*-.71*	.02	.22
CRS-4: "Academic Motivation"	.09	*.82*	.22	.15
CRS-2: "Aggressiveness"	.03	-.39	.18	*.68*
CI-1: "Attitude Toward Teacher"	-.06	.31	-.34	*.52*
% of Total Variance Accounted For	16.4	18.5	16.2	8.6

[1]Principal components solution, varimax rotation.

Relationship of Measures to "Social Competence"

As in analyses of spring 1977 and 1978 data, the relationships of test scores to a set of measures established as proxy criteria for "social competence" were examined. The social competence criteria consisted of the two subscales of the POCL ("Task Orientation" and "Sociability") and the four subscales of the Child Rating Scale ("Self-Assurance," "Aggressiveness," "Dependence" and "Academic Motivation"). These measures were selected in Phase I as indices of social competence because they appeared to tap a broad range of characteristics--social, emotional, cognitive, linguistic, and psychomotor--that contribute to children's everyday effectiveness as judged, in this instance, by their teachers and testers. Given the state-of-the-art of conceptualization and measurement of social competence, the criterion measures selected can provide, of course, only a crude estimate of the construct of interest.

The analytic design involved partitioning the variance in spring 1979 test scores into two parts: first, that part explained by demographic characteristics of the children (age, sex, and ethnicity); second, any additional variance accounted for by the complete set of social competence indices. It is hypothesized that the stronger the relationships of test scores to indices of social competence (over and above demographic factors), the more relevant these scores are as indices of program impact.

All of the test scores were found to be significantly related to the set of social competence criteria over and above any relationship to demographic factors, as can be seen in Table 19. The strongest relationship was found for the PIAT-Reading (22% of variance explained); the weakest, for the BSM-English (only 5% of variance explained).

Table 19

Relationship of "Social Competence" Criteria and Background Variables to Test Scores of First Grade Children

Test (N = 437)	% of variance accounted for jointly by "social competence" criteria & background variables	% of variance accounted for by "social competence" criteria beyond background variables	Significant predictors ($p \leq .05$) & their partial correlations with test scores	
			Variable	Partial
BSM-English	.19*	.05*	Ethnicity	-.34
			POCL-1	.16
			CRS-1	.13
Verbal Fluency	.11*	.11*	POCL-1	.16
			CRS-2	.16
PIAT-Math	.29*	.14*	Ethnicity	-.34
			POCL-1	.25
			Age	.18
			CRS-3	-.13
			POCL-2	-.11
PIAT-Reading	.27*	.22*	POCL-1	.24
			Ethnicity	-.17
			CRS-3	-.13

*Probability of associated F ratio $\leq$ .0002.

KEY:
POCL-1: "Task Orientation"
POCL-2: "Sociability"
CRS-1: "Self-Assurance"
CRS-2: "Aggressiveness"
CRS-3: "Dependence"
CRS-4: "Academic Motivation"

V

EXAMINATION OF OUTCOMES FOR CHILDREN

This chapter focuses on analyses dealing with child-level outcome measures for grade 1. It includes three major sections, dealing successively with three research questions:

- What are the impacts of PDC on children at the end of grade 1?
- What is the relation of measures in other study domains to child outcomes, irrespective of the educational treatment?
- To what extent do study measures in other domains help explain treatment-related differences in child outcomes?

Because of the length and complexity of the answers to these questions, a summary of the findings of analyses related to these questions appears separately as Chapter VI of this volume.

Evidence of PDC's Impact on Children

Guiding Questions

There are many aspects to the question "What has PDC's impact on children been?" These differing aspects imply various strategies in analysis and interpretation of project data. The purpose of these strategies is to set priorities in answering questions, to isolate potentially confounding factors, and to assess the merits of different responses to the main question. The aspects considered in designing an analytic approach to the PDC child outcome data have included:

- For what measures of child outcome, or for what aggregates or constructs of different measures, are effects to be assessed?
- Has the program's impact differed by site?
- What have been the effects of sample attrition on the comparability of PDC and comparison samples, and what effect does this have on outcome assessments?
- Have program impacts differed for children of differing demographic or background characteristics?
- How have program impacts differed over time?

Overview of the Child Impact Analytic Design

The basic analytic design for assessment of child impacts employs both univariate and multivariate analyses of variance and covariance.

In answering the main question, "What is the impact of PDC on children to date?", five specific questions are posed. The first two of these are understood as preliminary to the last three, which are roughly parallel to one another and explore different aspects of the answer to the questions. As will be shown below, the last three questions also refer to somewhat different samples of children. Figure 16 relates the five questions regarding child impacts to a sequence of decisions about analyses and describes the specific analytic design used in addressing each. The five questions are:

1. *For the children tested in spring 1979, are groups comparable on entry-level (fall 1976) data?*

2. *Is there a relation between child background characteristics and program effects--in particular, are there program-by-background interactions?*

3. *Looking at child outcome variables for spring 1979 only and considering outcome variables one at a time, are there effects of program, or program-by-site interactions?*

4. *Looking at child outcome variables for spring 1979 only and considering all of them simultaneously, are there program effects or program-by-site interactions?*

5. *Considering all possible occasions for each child outcome variable measured in spring 1979, are there trends and patterns over time in program or program-by-site interactions?*

Variables used in child outcome analyses. The following sets of variables were incorporated in one or more of the analyses discussed in this section:

Independent variables:

program conditions (two)
sites (eleven)

Child background characteristics:

sex (two levels)
prior preschool experience (two levels)
ethnicity (five levels)

Figure 16

Sequence of Research Questions and Analyses in the Examination of Child-Level Impacts at Grade 1 and Earlier Levels

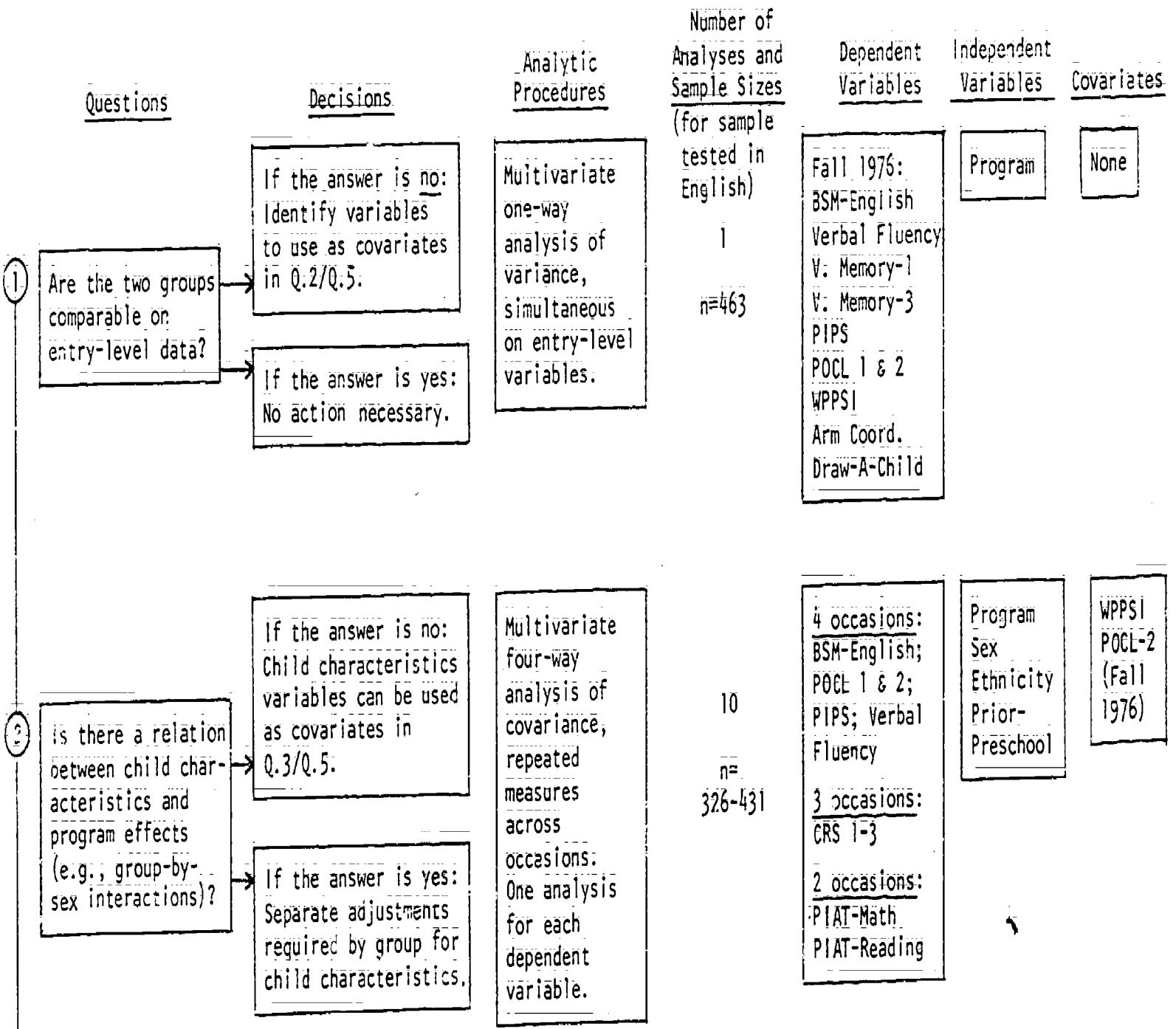


Figure 16
(continued)

	Questions	Analytic Procedures	Number of Analyses and Sample Sizes (for sample tested in English)	Dependent Variables	Independent Variables	Covariates (for sample tested in English)
B						
3	For spring 1979 outcomes (taken one at a time), are there group or group-by-site effects?	Univariate two-way analyses of covariance. One analysis for each dependent variable.	13 n= 452-480	BSM-English Verbal Fluency PIAT-Math PIAT-Reading PIPS POCL 1 & 2 CI 1 & 2 CRS 1-4	Program Site Dominant Language	Ethnicity Sex Prior Preschool Fall 1976 WPPSI POCL-2
4	For spring 1979 outcomes (taken simultaneously), are there group or group-by-site effects?	Multivariate two-way analysis of covariance.	1 n=376	Variables as for Q.3	Variables as for Q.3	Variables as for Q.3
5	For all occasions: are those patterns over time in group or group-by-site effects?	Multivariate two-way analysis of covariance. One analysis for each dependent variable across occasions.	13 n= 326-431	Variables as for Q.2	Variables as for Q.2	Variables as for Q.2

Entry-level variables:

BSM-English
BSM-Spanish
WPPSI
Verbal Fluency
Verbal Memory 1
Verbal Memory 3
PIPS
Draw-a-Child
Arm Coordination
POCL-1
POCL-2

Dependent variables:

BSM-English (four occasions)
BSM-Spanish (four occasions)
Verbal Fluency (four occasions)
PIPS (four occasions)
POCL-1 (four occasions)
POCL-2 (four occasions)
CRS-1 (three occasions)
CRS-2 (three occasions)
CRS-3 (three occasions)
PIAT-Math (two occasions)
PIAT-Reading (two occasions)
CRS-4 (one occasion)
Child Interview 1 (one occasion)
Child Interview 2 (one occasion)

Preliminary Analyses: Questions 1 and 2

In order to formulate appropriate analytic designs for assessing child outcomes in response to questions 3 through 5, it was first necessary to address questions 1 and 2.

Question 1: *Were PDC and comparison children remaining in the sample at the end of first grade (spring 1979) comparable at entry (fall 1976)?*

Findings. The methods used to address this question and the findings have already been described in detail in Chapter [illegible] (see in particular Table 10). We reiterate the findings here in summary form:

- For the subsample tested at entry in English (n=480), significant differences between PDC and comparison children were found for two of ten entry-level measures: WPPSI and POCL-2. Absolute differences between group means favored the comparison group in nine of ten instances. Multivariate analysis of variance considering all ten entry-level measures simultaneously confirmed the univariate ANOVA findings.

- For the subsample tested at entry in Spanish (n=71), no significant entry-level differences between PDC and comparison groups were found in either univariate or multivariate analyses of variance.

Analytic design implications. The findings for question 1 had important implications for the analyses that followed. First, the discovery of entry-level differences between PDC and comparison children tested in English meant that spring 1979 outcome comparisons were likely to be biased in favor of the comparison group unless outcome variable variance resulting from initial differences could be removed. Therefore, a decision was made to covary outcome variables on entry-level WPPSI and POCL-2 scores (the only two entry-level variables significantly differentiating groups) in all subsequent analyses for the sample initially tested in English. The adequacy of the analysis of covariance approach taken here to compensate for initial PDC/comparison differences will be considered in a later section, where we interpret statistical findings related to questions 3 through 5. Second, the decision to make covariance adjustments of outcome measures for the sample initially tested in English made it impractical, if not impossible, to analyze data for children initially tested in Spanish together with data for those tested in English. Because Spanish and English versions of entry-level measures cannot be assumed to measure the same constructs, analyses of the combined samples would have required separate covariate adjustments for children in the two groups--making analytic designs even more complicated and raising additional questions about interpretability. Consequently, it was decided that all analyses of outcomes would be performed separately for children tested initially in Spanish and in English.

Question 2: *Is there a relationship between child background characteristics and program effects--i.e., are there group-by-background interactions?*

Method of analysis. Each outcome variable was considered separately in repeated measures analyses incorporating all occasions of measurement of that variable from fall 1976 through spring 1979. The repeated measures analyses were performed using multivariate, rather than univariate, analysis of variance techniques in order to circumvent assumptions of compound symmetry (Winer, 1971). The analytic design incorporated main effects of group, sex, prior preschool experience, and ethnicity (English-tested sample only); separate first-order interactions of group with sex, prior preschool, and ethnicity (English-tested sample only); pooled higher-order interactions; and entry-level WPPSI and POCL-2 as covariates for the subsample initially tested in English. The number of measurements available ranged from one to four depending upon the variable; for measures administered only once, univariate approaches were used. Analytic sample sizes differed somewhat across variables due to missing data.

Findings. Table 20 summarizes the tests of first-order interactions for the sample initially tested in English. Pooled higher-order interactions were tested but not interpreted because of their complexity and because they are of little a priori interest at this stage of data analysis. A number of first-order interactions reached statistical significance; however, no consistent pattern emerged. And when the three significant group-by-ethnicity interactions were compared within the multivariate repeated measures analysis of variance framework, they were found to have rather different characteristics. When analyses were performed on data for children initially tested in Spanish, no significant first-order interactions were found.

Analytic design implications. Prior to addressing question 2 it was decided that interaction terms should not be incorporated in analytic designs for questions 3 through 5 unless there were compelling reasons to do so, because to include these terms as independent variables would entail estimating variation for an extremely large number of design cells, most of which would have few or no subjects. Under these circumstances it would be impossible to obtain consistently reliable estimates of within-cell variance, and the interpretability of the entire analysis would be jeopardized.

Tests of first-order interactions of group membership with background characteristics did not reveal strong, consistent interaction effects. Thus, we concluded that these interactions could be disregarded in subsequent analyses and that the background characteristics of sex, prior preschool experience, and ethnicity could be incorporated as covariates rather than design factors.

Analyses of Child Outcomes: Questions 3, 4, and 5

The analytic designs used to address questions 3 through 5 all involve adjustments of dependent variables for entry-level WPPSI and POCL-2 differences between groups (in the case of children initially tested in English) and for the main effects of sex, prior preschool experience, and (in the case of the sample tested in English) ethnicity. Because tests of the group main effect and the group-by-site interaction are performed on adjusted rather than observed scores, it is useful to examine mean adjusted scores for each dependent variable as well as the observed score means. Figures 17a through 18d graph the adjusted means for each occasion of measurement for each dependent variable by program group and language of initial testing. Tables 21 and 22 present both adjusted and observed mean values for each variable at each timepoint by group and language of initial testing.

The data presented in Figures 17a through 18d and Tables 21 and 22 apply strictly to the samples of children for whom scores were obtained at all occasions of measurement of the variable in question. These are the samples analyzed in addressing question 5. The relationships between these samples and the samples of analyses addressing questions 3 and 4 are illustrated in Table 23, which lists the sample sizes for each analysis and dependent variable.

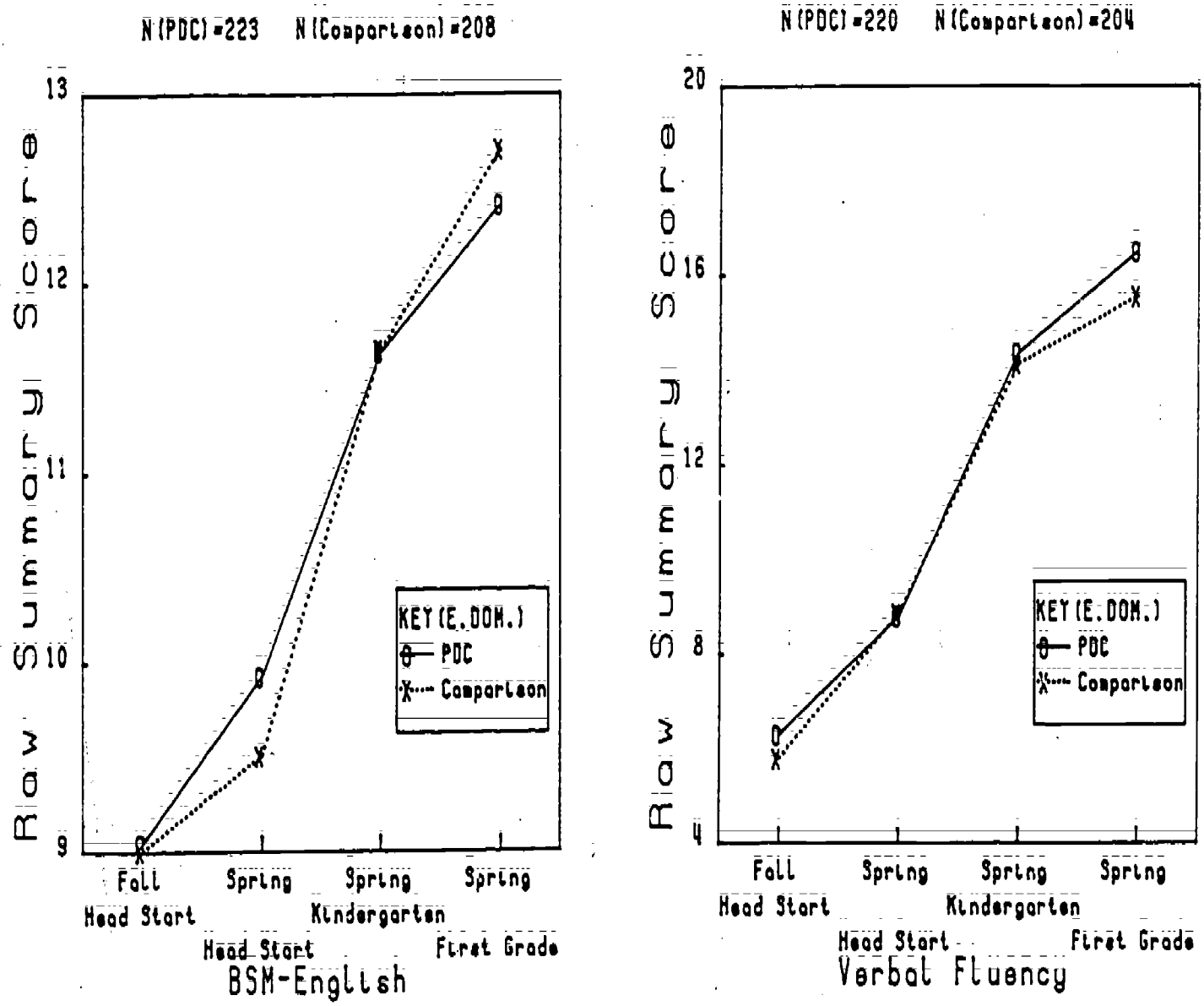
Table 20

Question 2 Findings: Summary of Tests of Interactions of Group with Ethnicity, Sex, and Prior Preschool Experience for Sample Tested in English

Interactions	Dependent Variables											
	BSM-English n=431	PIPS n=424	POCL-1 n=356	POCL-2 n=356	Verbal Fluenc n=42	CRS-2 n=356	CRS-3 n=359	CRS-4 n=452	PIAT-Math n=424	PIAT-Read n=398	CI-1 n=478	CI-2 n=460
Group-by-Ethnicity	p<.05				p<.0					p<.05		
Group-by-Sex		p<.05										
Group-by-Prior Preschool										p<.05		

Figure 17a

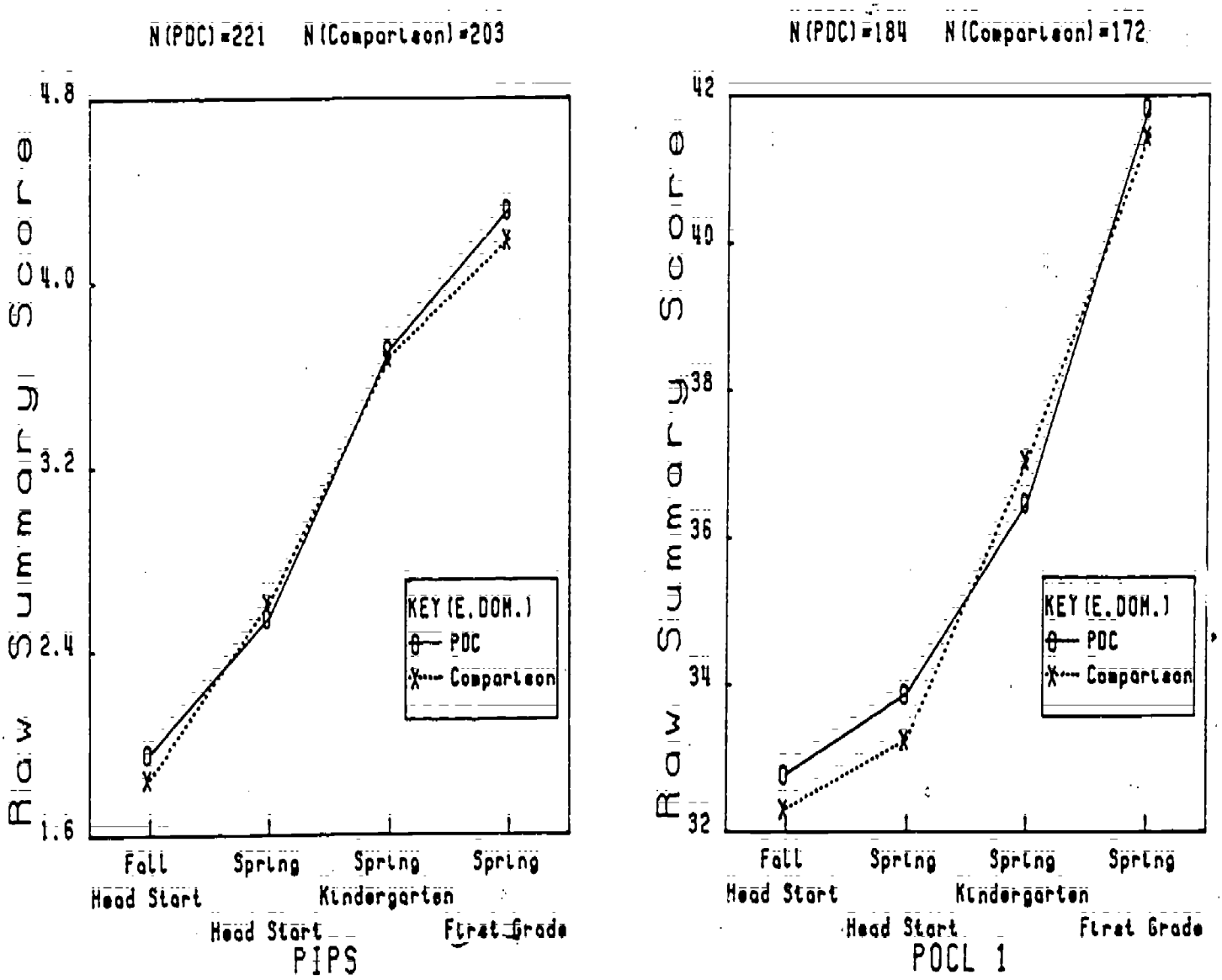
Plot of Adjusted PDC and Comparison Group Means on Measures in the Fall Head Start, Spring Kindergarten and Spring Grade 1 Batteries
(Sample tested in English)



Note: Each of the means plotted is adjusted for the effects of entry-level differences and differences in background characteristics. Residuals are readjusted to raw summary score value ranges by adding back overall sample means. Cases used for these plots are those with complete data for all timepoints.

Figure 17b

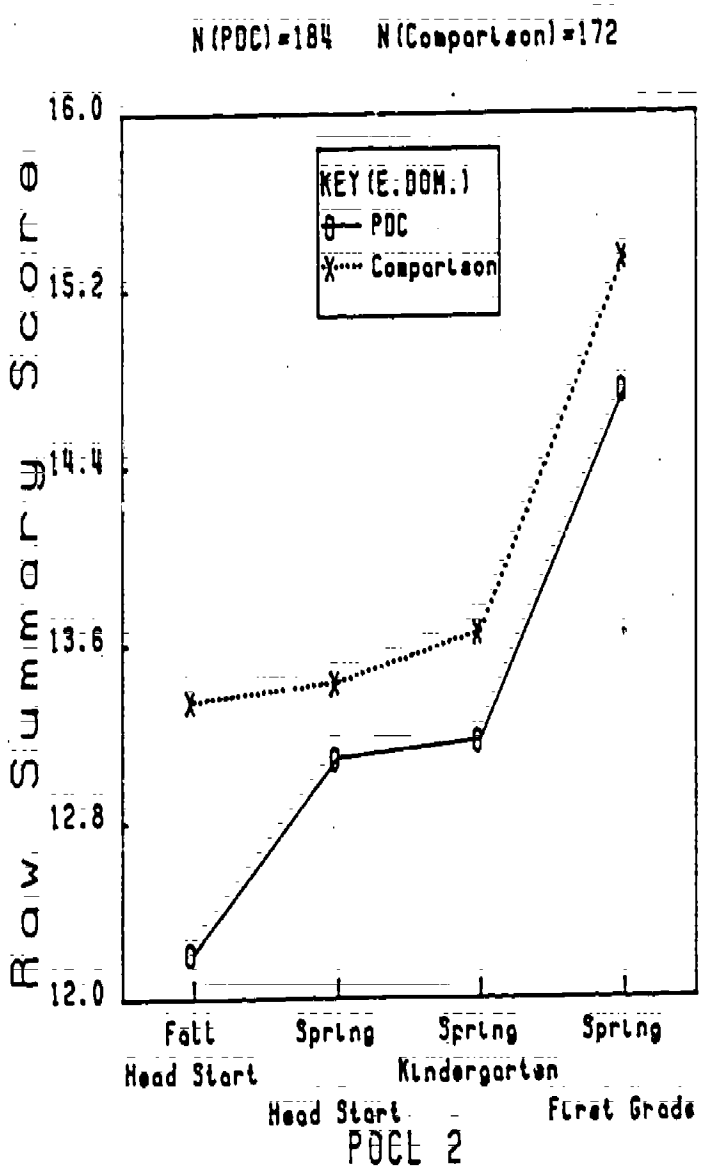
Plot of Adjusted PDC and Comparison Group Means on Measures in the Fall Head Start, Spring Kindergarten and Spring Grade 1 Batteries
(Sample tested in English)

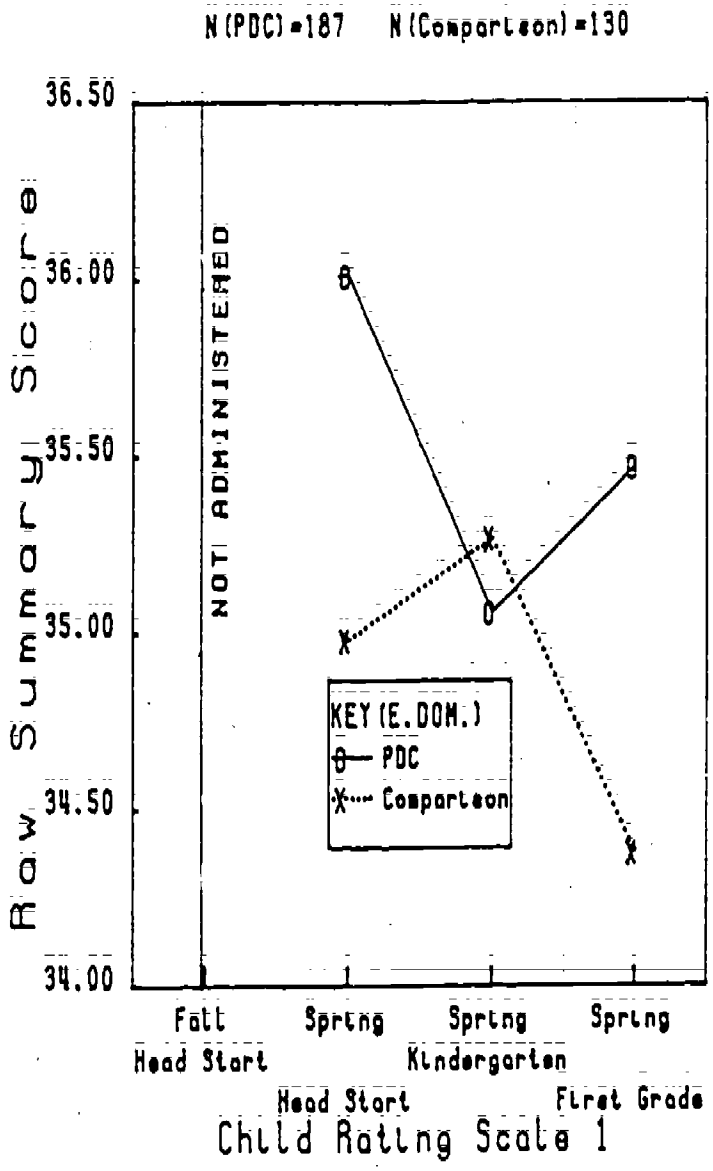


Note: Each of the means plotted is adjusted for the effects of entry-level differences and differences in background characteristics. Residuals are readjusted to raw summary score value ranges by adding back overall sample means. Cases used for these plots are those with complete data for all timepoints.

Figure 17c

Plot of Adjusted PDC and Comparison Group Means on Measures in the Fall Head Start, Spring Kindergarten and Spring Grade 1 Batteries
(Sample tested in English)

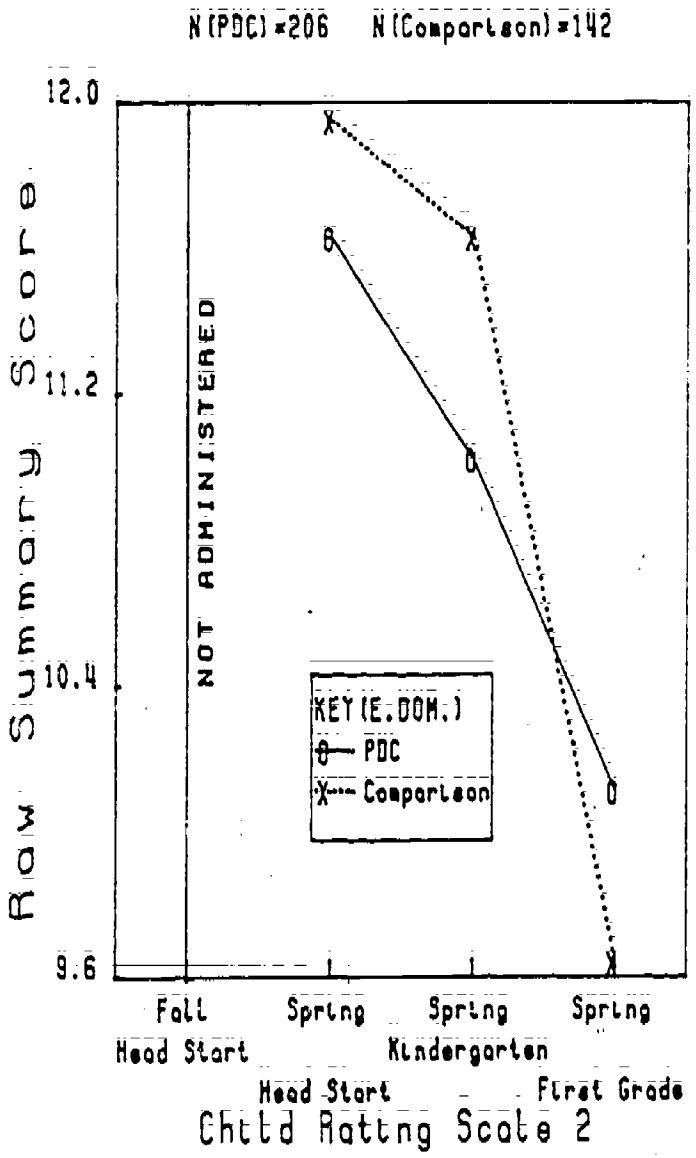


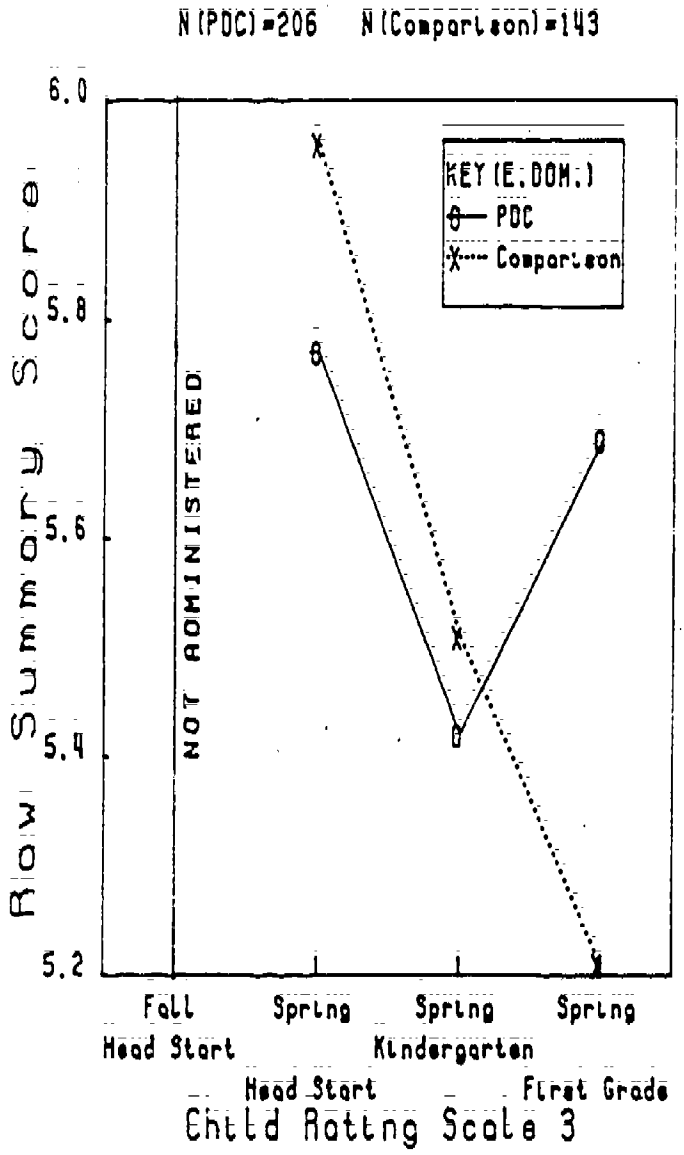


Note: Each of the means plotted is adjusted for the effects of entry-level differences and differences in background characteristics. Residuals are readjusted to raw summary score value ranges by adding back overall sample means. Cases used for these plots are those with complete data for all timepoints.

Figure 17d

Plot of Adjusted PDC and Comparison Group Means on Measures in the Fall Head Start, Spring Kindergarten and Spring Grade 1 Batteries
(Sample tested in English)

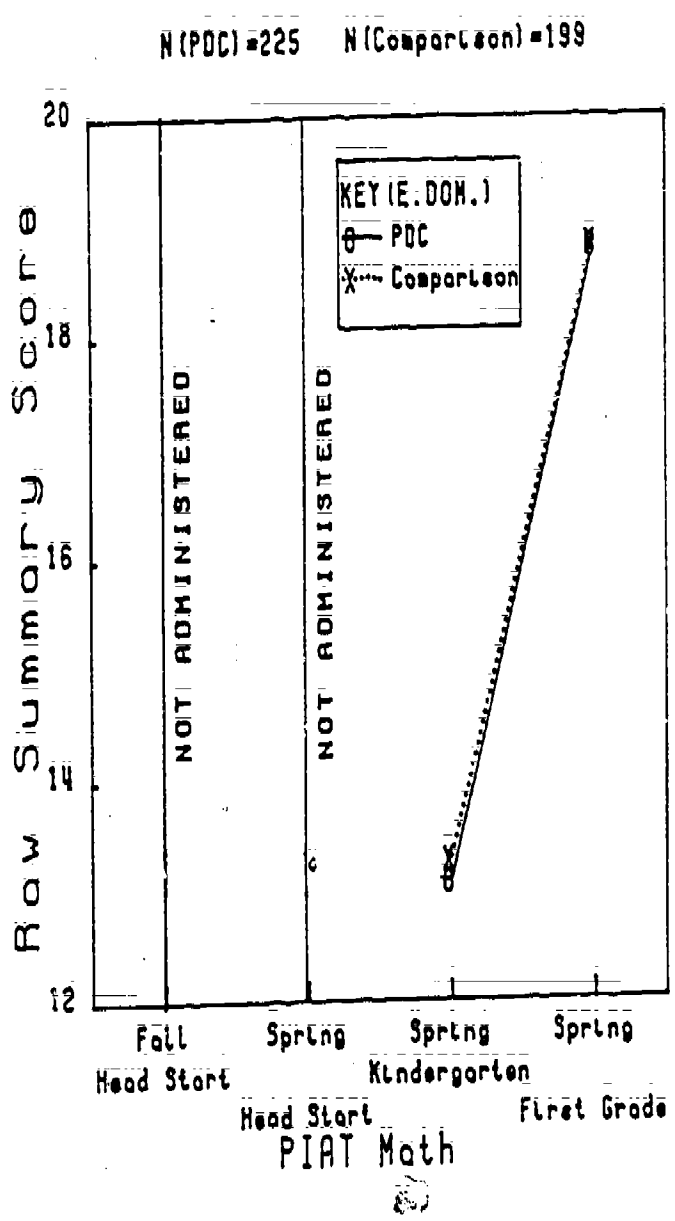


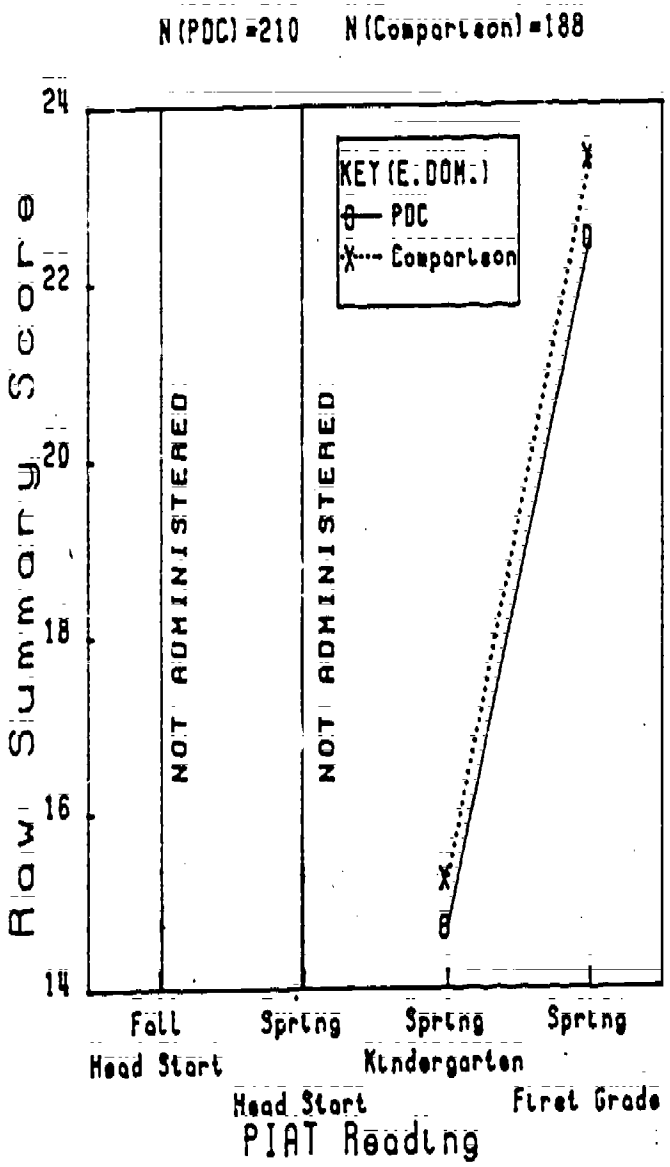


Note: Each of the means plotted is adjusted for the effects of entry-level differences and differences in background characteristics. Residuals are readjusted to raw summary score value ranges by adding back overall sample means. Cases used for these plots are those with complete data for all timepoints.

Figure 17e

Plot of Adjusted PDC and Comparison Group Means on Measures in the Fall Head Start, Spring Kindergarten and Spring Grade 1 Batteries
(Sample tested in English)

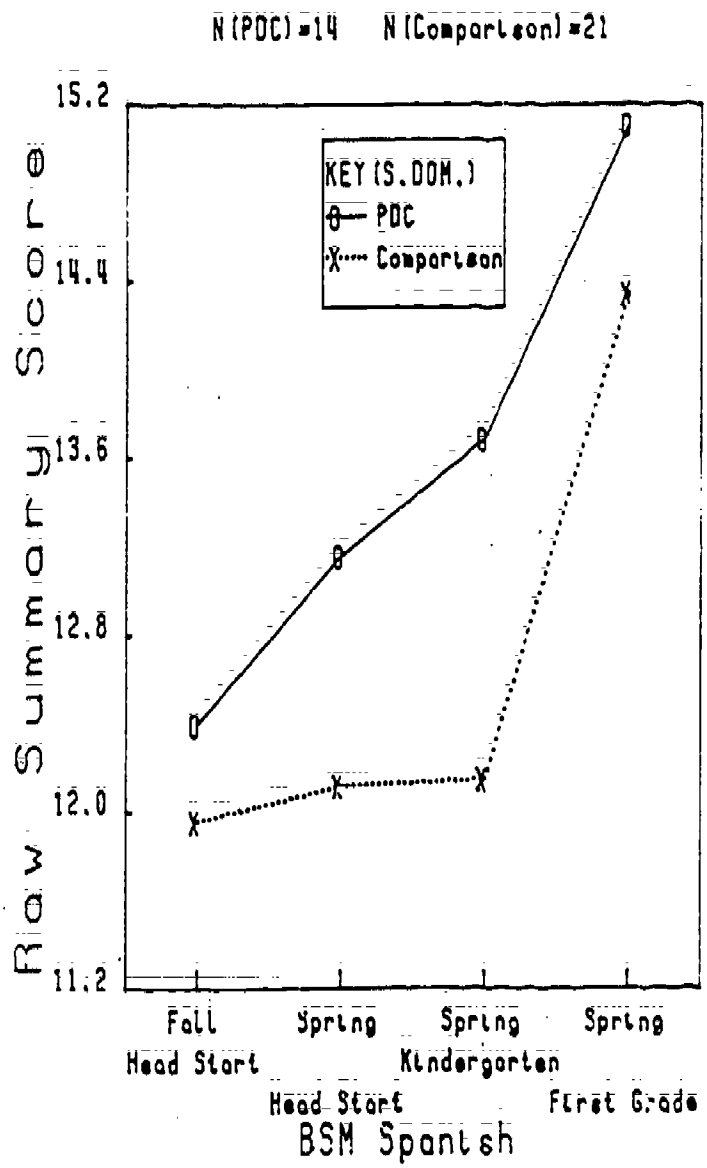


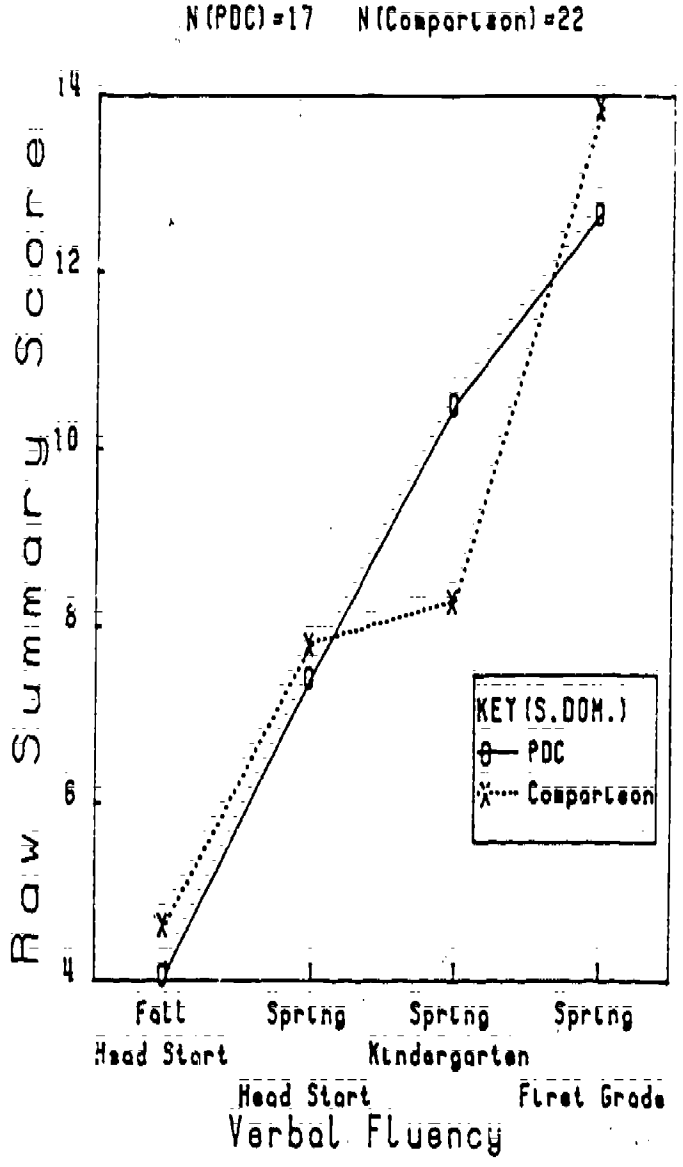


Note: Each of the means plotted is adjusted for the effects of entry-level differences and differences in background characteristics. Residuals are readjusted to raw summary score value ranges by adding back overall sample means. Cases used for these plots are those with complete data for all timepoints.

Figure 18a

Plot of Adjusted PDC and Comparison Group Means on Measures in the Fall Head Start, Spring Head Start, Spring Kindergarten and Spring Grade 1 Batteries (Sample tested in Spanish at Entry)

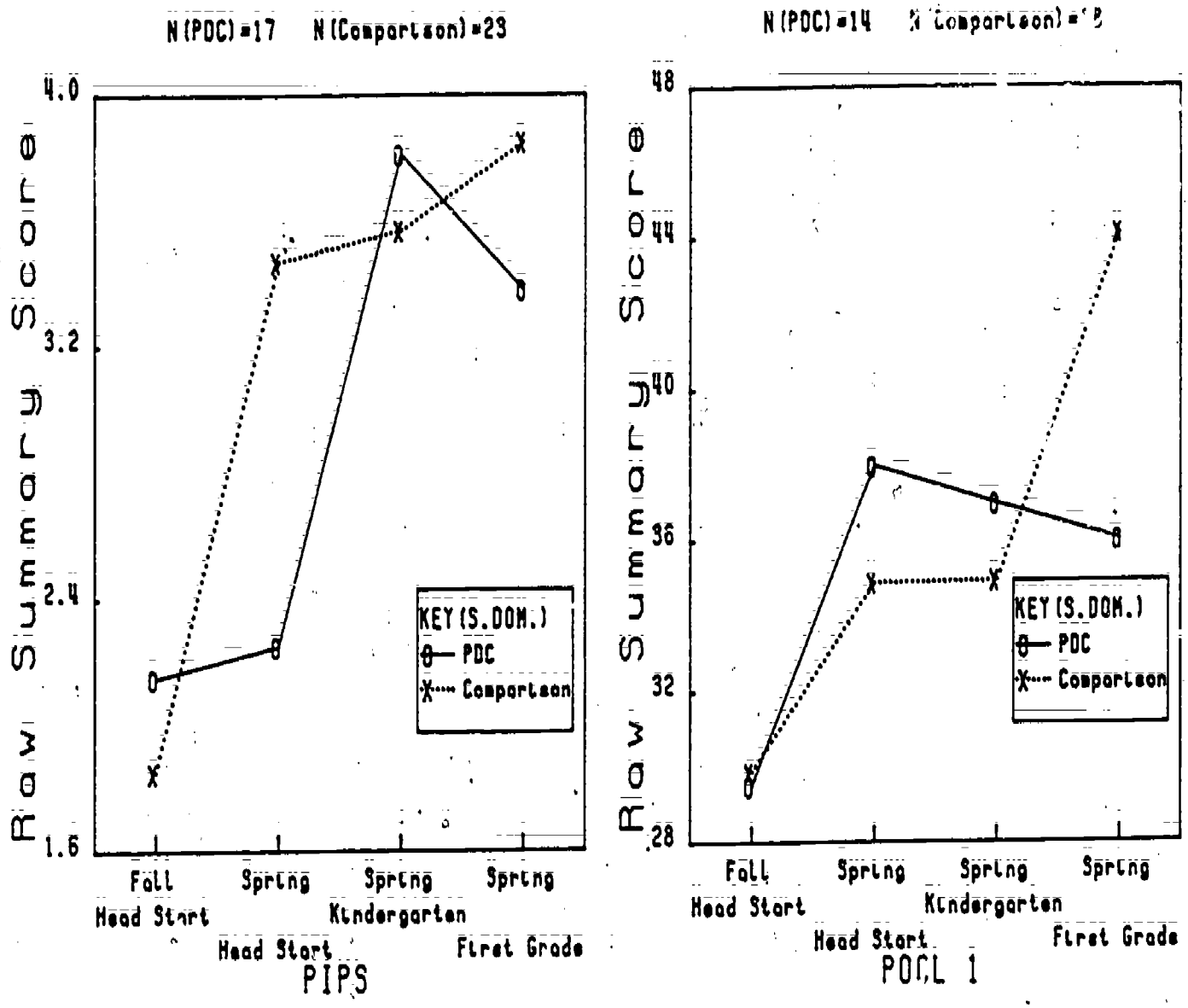




Note: Each of the means plotted is adjusted for the effects of differences in background characteristics. Residuals are readjusted to raw summary score value ranges by adding back overall sample means. Cases used for these plots are those with complete data for all timepoints. For the CRS-1, CRS-2, CRS-3 and BSM-English instruments, insufficient samples are available over time for analyses.

Figure 18b

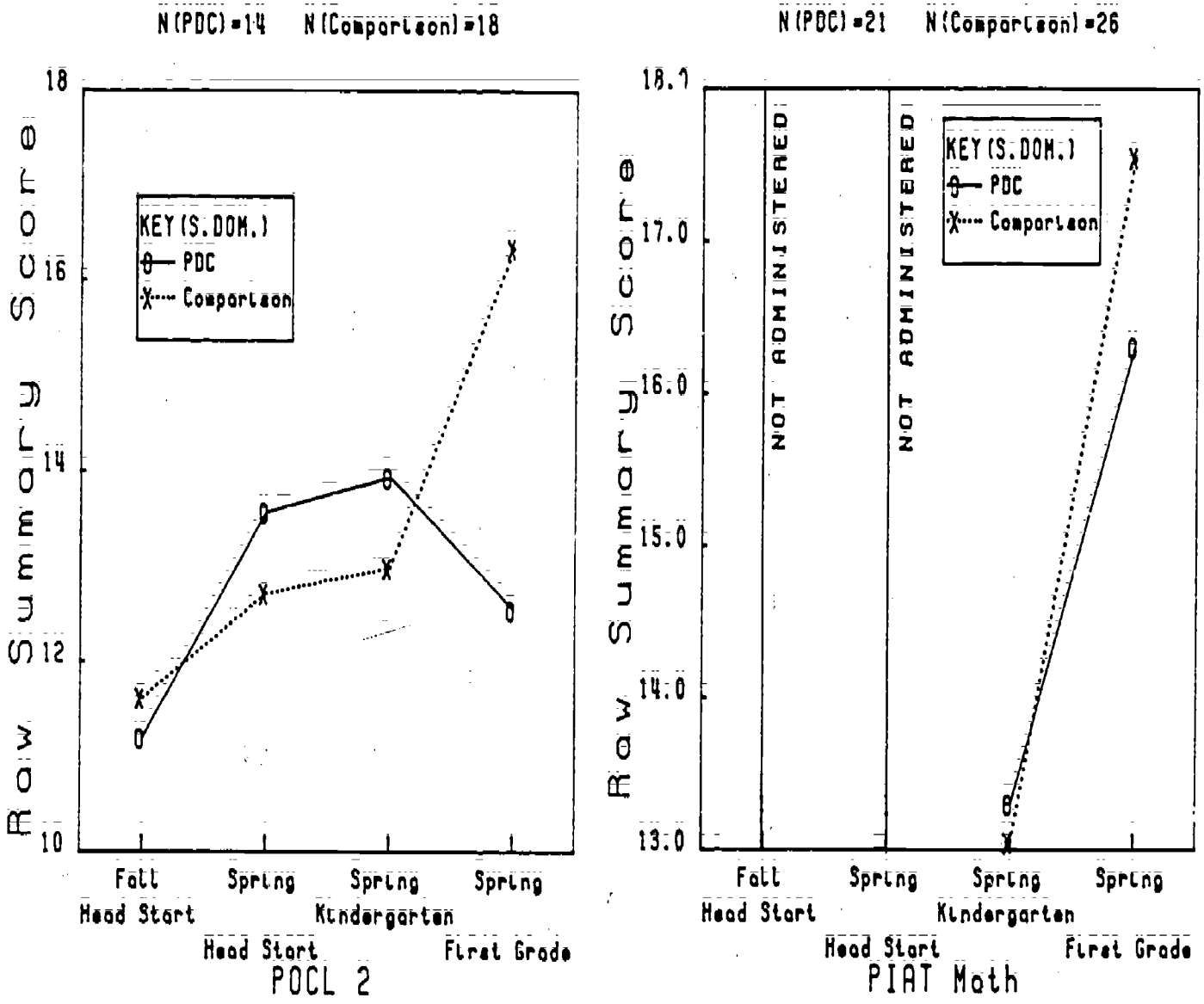
Plot of Adjusted PDC and Comparison Group Means on Measures in the Fall Head Start, Spring Head Start, Spring Kindergarten and Spring Grade 1 Batteries (Sample tested in Spanish at Entry)



Note: Each of the means plotted is adjusted for the effects of differences in background characteristics. Residuals are readjusted to raw summary score value ranges by adding back overall sample means. Cases used for these plots are those with complete data for all timepoints. For the CRS-1, CRS-2, CRS-3 and BSM-English instruments, insufficient samples are available over time for analyses.

Figure 18c

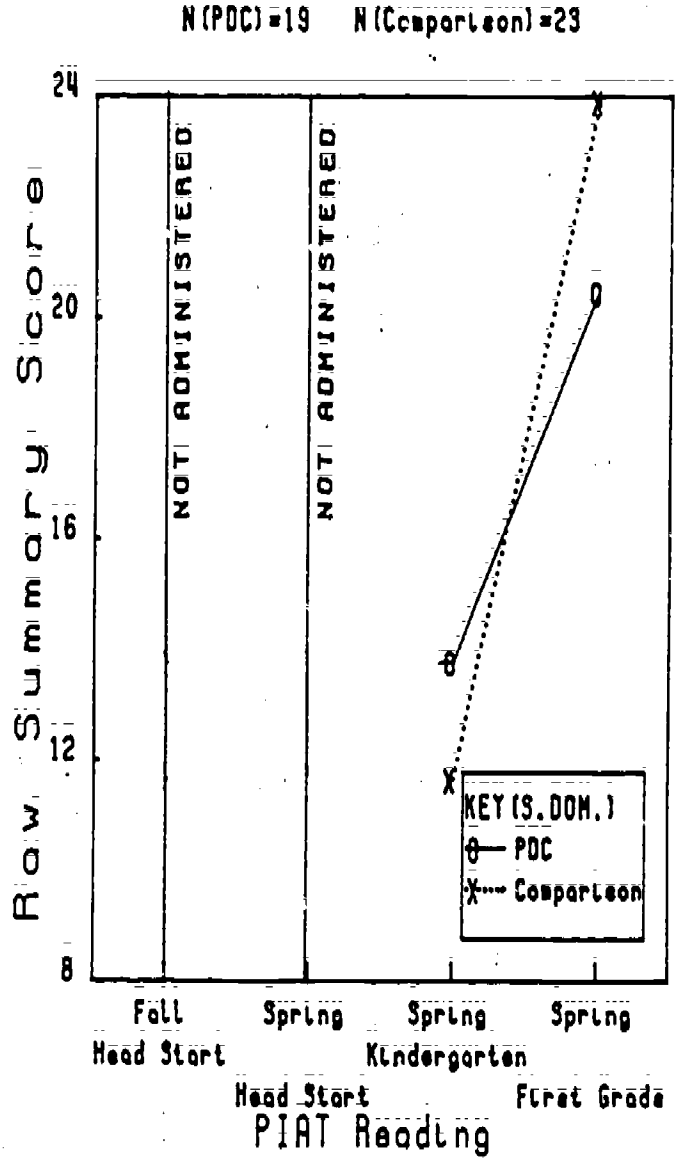
Plot of Adjusted PDC and Comparison Group Means on Measures in the Fall Head Start, Spring Head Start, Spring Kindergarten and Spring Grade 1 Batteries
(Sample tested in Spanish at Entry)



Note: Each of the means plotted is adjusted for the effects of differences in background characteristics. Residuals are readjusted to raw summary score value ranges by adding back overall sample means. Cases used for these plots are those with complete data for all timepoints. For the CRS-1, CRS-2, CRS-3 and BSM-English instruments, insufficient samples are available over time for analyses.

Figure 18d

Plot of Adjusted PDC and Comparison Group Means on Measures in the Fall Head Start, Spring Head Start, Spring Kindergarten and Spring Grade 1 Batteries (Sample tested in Spanish at Entry)



Note: Each of the means plotted is adjusted for the effects of differences in background characteristics. Residuals are readjusted to raw summary score value ranges by adding back overall sample means. Cases used for these plots are those with complete data for all timepoints. For the CRS-1, CRS-2, CRS-3 and BSM-English instruments, insufficient samples are available over time for analyses.

Table 21

Observed and Adjusted[1] Means for PDC and Comparison Groups
By Occasion: All Child Outcome Measures;
Sample Tested Initially in English

Variable	Group									
	PDC					Comparison				
	Occasion					Occasion				
	1 FHS	2 SHS	3 SK	4 SG1	(n)	1 FHS	2 SHS	3 SK	4 SG1	(n)
BSM-English										
Observed	8.77	9.65	11.46	12.29	(223)	9.31	9.81	11.84	12.85	(208)
Adjusted	9.05	9.93	11.64	12.42		9.01	9.51	11.65	12.71	
PIPS										
Observed	1.86	2.47	3.69	4.29	(221)	1.97	2.70	3.70	4.24	(203)
Adjusted	1.96	2.55	3.71	4.32		1.85	2.61	3.68	4.19	
POCL-1										
Observed	31.60	33.14	36.10	41.40	(184)	33.62	34.06	37.48	41.95	(172)
Adjusted	32.80	33.89	36.49	41.85		32.33	33.27	37.07	41.47	
POCL-2										
Observed	12.11	13.00	13.06	14.67	(184)	13.46	13.52	13.77	15.45	(172)
Adjusted	12.21	13.09	13.17	14.75		13.35	13.43	13.66	15.35	
Verbal Fluency										
Observed	5.95	8.45	14.11	16.38	(220)	6.18	9.27	14.47	16.74	(204)
Adjusted	6.31	8.82	14.40	16.53		5.89	8.87	14.16	16.58	
CRS-1										
Observed	-----	35.91	35.02	35.44	(196)	-----	35.11	35.33	34.39	(130)
Adjusted	-----	36.02	35.06	35.47			34.98	35.27	34.38	
CRS-2										
Observed	-----	11.48	10.86	10.04	(213)	-----	12.18	11.87	9.73	(143)
Adjusted	-----	11.63	11.02	10.11		-----	11.95	11.63	9.64	
CRS-3										
Observed	-----	5.80	5.42	5.69	(215)	-----	5.92	5.51	5.20	(144)
Adjusted	-----	5.77	5.42	5.69		-----	5.96	5.51	5.21	

[1]Adjusted means are obtained as residuals of regression equations for each child outcome as dependent variable, using as predictors entry-level values for WPPSI and POCL-2, as well as dummy variables for sex, prior preschool experience and ethnicity. Residuals are adjusted by the addition of overall means for each occasion.

The sample represented in these data is the sample of children with complete data for all occasions of measurement for a given instrument.

Table 21
(continued)

Variable	Group									
	PDC					Comparison				
	Occasion					Occasion				
	1 FHS	2 SHS	3 SK	4 SG1	(n)	1 FHS	2 SHS	3 SK	4 SG1	(n)
PIAT-Math										
Observed	-----	-----	12.88	18.62	(225)	-----	-----	13.53	19.56	(199)
Adjusted	-----	-----	13.09	18.83		-----	-----	13.27	18.86	
PIAT-Reading										
Observed	-----	-----	14.70	22.18	(210)	-----	-----	15.79	23.89	(188)
Adjusted	-----	-----	14.71	22.50		-----	-----	15.26	23.41	
CRS-4										
Observed	-----	-----	-----	9.93	(230)	-----	-----	-----	9.57	(222)
Adjusted	-----	-----	-----	10.00		-----	-----	-----	9.50	
Child Interview-1										
Observed	-----	-----	-----	2.29	(245)	-----	-----	-----	2.33	(233)
Adjusted	-----	-----	-----	2.28		-----	-----	-----	2.32	
Child Interview-2										
Observed	-----	-----	-----	9.64	(235)	-----	-----	-----	9.83	(225)
Adjusted	-----	-----	-----	9.69		-----	-----	-----	9.77	

Table 22

Observed and Adjusted[1] Means for PDC and Comparison Groups
By Occasion: All Child Outcome Measures;
Sample Tested Initially in Spanish

Variable	Group									
	PDC					Comparison				
	Occasion					Occasion				
	1 FHS	2 SHS	3 SK	4 SG1	(n)	1 FHS	2 SHS	3 SK	4 SG1	(n)
BSM-Spanish										
Observed	12.50	13.21	13.79	15.07	(14)	11.90	12.10	12.10	14.38	(21)
Adjusted	12.40	13.16	13.69	15.11		11.96	12.12	12.15	14.35	
PIPS										
Observed	2.12	2.29	3.76	3.18	(17)	1.97	3.43	3.61	4.00	(23)
Adjusted	2.15	2.25	3.81	3.38		1.85	3.47	3.57	3.85	
POCL-1										
Observed	30.93	39.14	37.21	36.43	(14)	28.78	34.00	34.78	43.83	(18)
Adjusted	29.52	37.99	36.99	36.02		29.88	34.90	34.96	44.15	
POCL-2										
Observed	11.64	13.71	14.07	13.21	(14)	11.28	12.61	12.89	15.83	(18)
Adjusted	11.20	13.57	13.94	12.55		11.62	12.72	12.99	16.35	
Verbal Fluency										
Observed	4.06	7.65	10.59	12.76	(17)	4.68	7.68	8.27	13.82	(22)
Adjusted	4.09	7.45	10.54	12.69		4.65	7.83	8.31	13.88	
CRS-1										
Observed	S A M P L E S I Z E T O O S M A L L F O R A N A L Y S I S									
Adjusted										
CRS-2										
Observed	S A M P L E S I Z E T O O S M A L L F O R A N A L Y S I S									
Adjusted										
CRS-3										
Observed	S A M P L E S I Z E T O O S M A L L F O R A N A L Y S I S									
Adjusted										

[1]Adjusted means are obtained as residuals of regression equations for each child outcome as dependent variable, using as predictors dummy variables for sex and prior preschool experience. Residuals are adjusted by the addition of overall means for each occasion.

The sample represented in these data is the sample of children with complete data for all measurement occasions for a given instrument.

Table 22
(continued)

Variable	Group									
	PDC					Comparison				
	Occasion					Occasion				
	1 FHS	2 SHS	3 SK	4 SGI	(n)	1 FHS	2 SHS	3 SK	4 SGI	(n)
PIAT-Math										
Observed	-----	-----	13.38	16.67	(21)	-----	-----	13.00	17.27	(26)
Adjusted	-----	-----	13.30	16.31		-----	-----	13.05	17.55	
PIAT-Reading										
Observed	-----	-----	13.90	20.53	(19)	-----	-----	11.56	23.87	(23)
Adjusted	-----	-----	13.78	20.47		-----	-----	11.65	23.91	
CRS-4										
Observed	-----	-----	-----	10.30	(37)	-----	-----	-----	8.79	(33)
Adjusted	-----	-----	-----	10.26		-----	-----	-----	8.83	
Child Interview-1										
Observed	-----	-----	-----	2.38	(37)	-----	-----	-----	2.11	(33)
Adjusted	-----	-----	-----	2.35		-----	-----	-----	2.14	
Child Interview-2										
Observed	-----	-----	-----	7.22	(37)	-----	-----	-----	7.32	(31)
Adjusted	-----	-----	-----	7.28		-----	-----	-----	7.25	

Table 23

Comparison of Sample Sizes Available for Analyses
Addressing Questions 3, 4 and 5, by Child Outcome Measure
for Sample Tested in English and (at Entry) in Spanish

Measure	Sample Tested in English n = 480			Sample Tested in Spanish n = 71		
	Questions			Questions		
	3	4	5	3	4	5
BSM-English	480	376	431	70	SI	SI
BSM-Spanish	SI	SI	SI	70	50	35
PIPS	479	376	424	70	50	40
POCL-1	454	376	356	62	50	32
POCL-2	455	376	356	62	50	32
Verbal Fluency	480	376	424	71	50	39
CRS-1	449	376	326	70	50	SI
CRS-2	458	376	356	68	50	SI
CRS-3	456	376	359	71	50	SI
CRS-4	452	376	---	70	50	--
PIAT-Math	471	376	424	68	50	47
PIAT-Reading	456	376	398	64	50	42
Child Interview-1	478	376	---	70	50	--
Child Interview-2	460	376	---	68	50	--

Notes: Analyses for question 3 use all available subjects with spring 1979 data (and values for entry-level covariates); for question 4, all available subjects with complete data across all instruments in spring 1979; and for question 5, all available subjects with complete data across all occasions for each instrument. The entry "SI" means that the sample of children with data for that measure is insufficient for analysis for a particular question. For single-occasion measures, question 3 and question 5 analyses are identical.

Question 3: *Considering outcome variables one at a time, are PDC and comparison children different at the end of first grade (spring 1979)? In particular, are there group main effects or group-by-site interactions?*

Method of analysis. This question was addressed by univariate two-way analyses of covariance. The dependent variables were each spring 1979 child outcome variable taken separately; the independent variables were group, site, and the group-by-site interaction term. Covariates were different for the samples initially tested in English and Spanish as previously indicated.

Findings for children initially tested in English. Table 24 summarizes the results of analyses for the subsample initially tested in English. A significant group main effect was found for PIAT-Reading; significant group-by-site interactions, for both PIAT-Reading and Verbal Fluency. Post hoc comparisons of PIAT-Reading adjusted mean scores for PDC and comparison groups within sites are presented in Table 25. These tests clearly indicate that the group-by-site interaction is the result of a simple effect of group within site--i.e., group membership is associated with a mean difference in PIAT-Reading scores favoring the comparison group at only one site, Michigan. Also evident from this table is the fact that Georgia, the only PDC site without a local comparison group, has the lowest mean PIAT-Reading scores of any site. Given the consistently strong site effect across all measures shown in Table 24, it is reasonable to assume that if there were a group of comparison children at the Georgia site, they would also have low mean scores on the PIAT-Reading. The inclusion of the Georgia PDC group without a comparison group serves to lower the PDC mean but not the comparison mean. As the bottom row of Table 25 shows, when the Georgia site is excluded from the analysis of PIAT-Reading, the overall difference in group means disappears. In sum, there is no basis for concluding that the overall performances of PDC and comparison groups initially tested in English are different on the PIAT-Reading subtest. Regarding the rather large difference in PIAT-Reading performance associated with group membership at the Michigan site, it is not clear that the effect should be attributed to differences in educational program.

Table 26 presents the results of post hoc comparisons of PDC and comparison group mean adjusted Verbal Fluency scores within each site. The significant group-by-site interaction for Verbal Fluency appears due to significant differences between groups going in opposite directions at two sites. In Maryland, children in the PDC group scored higher on Verbal Fluency than children in the comparison group; in Washington, comparison children scored higher than PDC children.

Findings for children initially tested in Spanish. Table 27 summarizes the results of univariate analyses of spring 1979 outcome variables for the PDC and comparison samples initially tested in Spanish. Significant differences associated with group membership appear for two variables--the CRS-4 "Academic Motivation" scale (PDC>Comp) and the PIAT-Reading (Comp>PDC). No significant group-by-site interactions were found.

Table 24

Results of Univariate Analyses of Covariance on Spring 1979 Child Outcome Measures: Sample Tested in English (maximum n=481)

Outcome Measure	N	EFFECTS			
		Covariates	Group	Site	Group-by-Site
BSM-English	480	.0001	--	.0001	--
PIPS	479	.0273	--	.0407	--
POCL-1	454	.0021	--	.0013	--
POCL-2	455	.0001	--	.0255	--
Verbal Fluency	480	.0019	--	.0001	.0017
CRS-1	449	.0003	--	.0081	--
CRS-2	458	.0036	--	.0013	--
CRS-3	456	--	--	.0024	--
CRS-4	452	.0007	--	.0126	--
PIAT-Math	471	.0001	--	.0475	--
PIAT-Reading	456	.0001	.0328	.0059	.0482
Child Interview 1	478	--	--	--	--
Child Interview 2	460	.0001	--	.0016	--

Note: Table entries are statistical significance levels, if lower than .05. Covariates for these analyses are: entry-level WPPSI and POCL-2, and dummy variables for sex, prior preschool experience and ethnicity.

Table 25

Adjusted Means for PDC and Comparison Groups at Each Site:
Spring 1979 PIAT-Reading Instrument (n=456); Sample Tested in English

Site	GROUP				Contrast
	PDC		Comparison		
	N	Mean	N	Mean	
California	17	20.58	17	20.63	NS
Colorado	20	24.24	12	23.58	NS
Connecticut	23	22.48	34	22.76	NS
Florida	28	21.53	25	24.46	NS
Georgia	27	20.27	--	--	--
Iowa	13	24.14	13	25.52	NS
Maryland	27	25.15	22	23.51	NS
Michigan	20	20.66	29	26.32	p<.0001
Texas	12	22.96	13	22.30	NS
Utah	22	23.74	32	23.32	NS
Washington	24	22.29	26	22.24	NS
All Sites	233	22.49	223	23.52	p=.0328
All Sites but Georgia	206	22.78	223	23.52	NS

Notes: Means are residuals of regression on observed data, using as independent variables entry-level WPPSI and POCL-2 scores and dummy variables for sex, ethnicity and prior preschool experience. Residuals are restored to overall mean levels. Significance levels for contrasts between treatment-group means are presented if .05 or less.

Table 26

Adjusted Means for PDC and Comparison Groups at Each Site:
Spring 1979 MSCA-Verbal Fluency Instrument (n=480); Sample Tested in English

Site	GROUP				Contrast
	PDC		Comparison		
	N	Mean	N	Mean	
California	17	16.02	18	16.76	NS
Colorado	20	11.27	14	13.98	NS
Connecticut	27	20.47	36	19.08	NS
Florida	31	13.08	26	13.79	NS
Georgia	27	13.82	--	--	--
Iowa	15	16.84	13	16.62	NS
Maryland	27	22.75	22	17.67	p=.0103
Michigan	21	15.82	32	16.39	NS
Texas	12	17.39	14	16.14	NS
Utah	23	17.02	34	15.52	NS
Washington	25	15.58	26	19.34	p=.0329
All Sites	245	16.40	235	16.72	NS

Notes: Means are residuals of regression on observed data, using as independent variables entry-level WPPSI and POCL-2 scores and dummy variables for sex, prior preschool experience and ethnicity. Residuals are restored to overall mean levels. Significance levels for contrasts between treatment group means are presented only if .05 or less.

Table 27

Results of Univariate Analyses of Covariance on Spring 1979 Child Outcome Measures: Sample Tested in Spanish (maximum n=70)

Outcome Measure	N	EFFECTS			
		Covariates	Group	Site	Group-by-Site
BSM-English	70	--	--	.0001	--
BSM-Spanish	70	--	--	--	--
PIPS	70	--	--	--	--
POCL-1	62	--	--	--	--
POCL-2	62	--	--	--	--
Verbal Fluency	70	--	--	--	--
CRS-1	70	--	--	--	--
CRS-2	68	--	--	--	--
CRS-3	70	--	--	.0362	--
CRS-4	70	--	.0131	.0001	--
PIAT-Math	68	--	--	--	--
PIAT-Reading	64	--	.0033	.0051	--
Child Interview 1	70	--	--	--	--
Child Interview 2	70	--	--	--	--

Note: Table entries are statistical significance levels, if lower than .05. Covariates for these analyses are dummy variables for sex and prior preschool experience.

Table 28 further explores the significant group main effects through post hoc within-site comparisons. For the Academic Motivation scale, PDC children were rated higher than comparison children only at the Texas site. By contrast, PIAT-Reading scores are significantly higher for comparison children at the Texas site. Although the direction and magnitude of difference in PIAT-Reading performance was similar at the California site, the sample sizes there were too small for the difference to reach statistical significance. Including data for six Connecticut PDC children (none of them in the comparison group) in analyses of the PIAT-Reading scores increased the magnitude of the group effect, much as the inclusion of data from the Georgia site affected the analysis of PIAT-Reading scores for the sample tested in English. However, in this instance, excluding the Connecticut sample did not eliminate the overall group difference.

Question 4: *Considering all outcome variables simultaneously, are PDC and comparison children different at the end of first grade (spring 1979)? In particular, is there a group main effect or a group-by-site interaction?*

Method of analysis. The overall group profiles of spring 1979 outcome variables were compared in a two-way multivariate analysis of covariance. The dependent variables were all spring 1979 outcomes considered simultaneously; independent variables were group, site, and the group-by-site interaction term. Covariates differed for children initially tested in English and in Spanish as previously indicated.

Relationship to analyses for question 3. The analyses performed in response to this question might be expected to produce somewhat different answers than those obtained to question 3 for two reasons. First, it is possible to find differences in overall outcome profiles in the presence of scattered and mixed univariate differences or even in the absence of any significant univariate differences. Second, the sample for the multivariate analyses reported in this section was smaller than for the univariate analyses in response to question 3, since some children in the spring 1979 sample were not tested or rated on some measures and only children with complete data were included in the multivariate analysis. Thus, multivariate analysis is not equivalent to the sum of univariate analyses but supplements them. If both approaches produce similar findings, these findings have greater credence. If the results contradict one another, an effort must be made to resolve the difference.

Findings for the sample initially tested in English. Table 29 presents the results of a multivariate analysis of covariance of spring 1979 outcome measures for the PDC and comparison samples initially tested in English. The total sample available was 376--78% of the 480 children in this subsample measured on at least one outcome variable in spring 1979. The top section of the table reveals a significant difference in the outcome profiles of children in the PDC and comparison groups as well as a significant group-by-site interaction. The bottom section of Table 29 indicates which variables made significant contributions to the overall difference in group profiles. The pattern of these results largely replicates findings of the univariate analyses for question 3, in spite of a reduction in sample size (and therefore, a difference in sample composition). Specific profile differences and their directions are summarized below:

Table 28

Adjusted Means for PDC and Comparison Groups at each Site for Spring 1979 (Grade 1) PIAT-Reading and CRS-4 "Academic Motivation" Scale, Sample Tested Over Time in Spanish

Variable	Site	Group				Contrast
		PDC		Comparison		
		n	Mean	n	Mean	
CRS-4	California	4	5.52	7	5.72	NS
	Connecticut	8	11.61	1	11.80	NS
	Texas	25	10.80	25	9.49	p=.0513
	All Sites	37	10.40	33	8.76	p=.0131
PIAT-Reading	California	4	16.89	7	21.50	NS
	Connecticut	6	15.39	--	-----	--
	Texas	24	21.54	23	24.08	p=.0402
	All Sites	34	19.91	30	23.48	p=.0033
	Sites, Connecticut Removed	28	20.88	30	23.48	p=.0248

Notes: Means are residuals of regression on observed data, using as independent variables dummy variables for sex and prior preschool experience. Residuals are restored to overall mean levels. Significance levels for contrasts between program group means are presented only if .05 or less.

Table 29

Results of Multivariate Analysis of Covariance of Spring 1979 (Grade 1) Child Outcome Measures: Sample Tested in English (n=376)

Outcome Measure	EFFECTS			
	Covariates	Group	Site	Group-by-Site
Overall Comparison of Profiles	.0001	.0009	.0001	.0056
Contributions of Each Outcome Measure				
BSM-English	.0001	.0289	.0001	--
PIPS	--	--	--	--
POCL-1	--	--	.0016	--
POCL-2	.0014	--	.0227	--
Verbal Fluency	.0018	--	.0001	.0011
CRS-1	.0005	--	.0197	--
CRS-2	.0031	--	.0010	--
CRS-3	--	.0389	.0010	--
CRS-4	.0001	.0280	.0250	--
PIAT-Math	.0001	--	--	--
PIAT Reading	.0001	.0184	.0090	.0477
Child Interview 1	--	--	--	--
Child Interview 2	.0013	--	.0342	--

Notes: Table entries are statistical significance levels, if lower than .05. Covariates for these analyses are dummy variables for sex, ethnicity and prior preschool experience, and entry-level values for WPPSI and POCL-2.

- PIAT-Reading. Differences in mean PIAT-Reading scores favored the comparison group. This difference is again interpreted as primarily the result of a large difference favoring the comparison group at the Michigan site and the absence of a comparison group in Georgia, a low scoring site. The difference between groups amounted to .19 standard deviation units.

- Verbal Fluency. A significant group-by-site interaction was found. As in the preceding univariate analyses, site-level comparisons point to significant differences in opposite directions at the Maryland and Washington sites, with no other site-level differences between groups.

- BSM-English. An overall difference favoring the comparison group was found: PDC=12.60; comparison=12.90. The difference amounted to only .13 standard deviations.

- CRS-3 "Dependence." There was an overall difference favoring comparison group children: PDC=5.67; comparison=5.47. (Higher scores indicate greater "dependence.") The magnitude of this difference was .11 standard deviations.

- CRS-4 "Academic Motivation." The overall difference favored PDC children: PDC=10.00; comparison=9.40. This difference amounted to .19 standard deviations.

Although the outcome profiles for PDC and comparison groups overall are reliably different, the implication of this difference is not clear (even if we leave aside the question of whether the difference can be attributed to the educational program rather than to other factors). First, neither group's profile is consistently higher than the other; second, what differences there are tend to be confined to particular measures at particular sites. Finally, the group differences are small (on the order of two-tenths of a standard deviation in magnitude), suggesting that the multivariate analytic procedures may be more sensitive than is justified by the meaning of the educational impacts which have been identified, given our overall large sample size and the strong influence of site characteristics.

Findings for children initially tested in Spanish. The same analysis was carried out for children initially tested in Spanish. The available sample size was 50--71% of the 70 children in this subsample measured on at least one outcome variable in spring 1979. Children were distributed very unevenly across sites. There were only seven children representing the California site, and only two of these belonged to the PDC group. There were only five children from Connecticut, all in the PDC group. The remaining 38 children were from Texas. Given gross imbalances in sample sizes across sites and groups and the excessive number of dependent variables for the number of children in the total sample, the analytic results were judged too unstable to report.

Question 5: *Considering all occasions of measurement of each child outcome variable from fall 1976 through spring 1979, are there trends and patterns over time in group differences or group-by-site interactions?*

Method of analysis. The repeated measures analysis for each instrument was accomplished by a multivariate two-way analysis of covariance. Ten of the 13 instruments considered in this section were administered on more than one occasion; three--CRS-4, CI-1, and CI-2--were first administered in spring 1979 and have not been included in the repeated measures analyses. The independent variables were group and site. Covariates remained the same as for previous analyses. For reasons analogous to those articulated in the previous section, analyses of data for children initially tested in Spanish were not conducted.

Findings for children initially tested in English. Table 30 summarizes the results of analyses for each outcome measure. Significant differences over time associated with group membership were identified for two variables: the POCL-2 "Sociability" scale and the CRS-2 "Aggressiveness" scale. Group-by-site interactions over time emerged for four measures: Verbal Fluency, CRS-2 "Aggressiveness" scale, CRS-3 "Dependence" scale, and PIAT-Reading. The nature of each of these longitudinal effects is considered below:

- POCL-2 main effect. When group mean total POCL-2 scores--the sum of values across four occasions--were compared, comparison children were found to score significantly higher than PDC children. This difference indicates that comparison children as a group were judged over time to be slightly more "sociable" than PDC children. The magnitude of the difference was approximately one-quarter of a standard deviation. Differences between groups did not appear for any single occasion; neither was there any indication of differences in trends or patterns over time. Visual inspection of POCL-2 curves in Figure 17c shows that the largest difference between the two groups occurred during the fall of their Head Start years before children had spent any significant amount of time in the PDC program.

- CRS-2 main effect and interaction. CRS-2 "Aggressiveness" scale ratings were made on three occasions beginning in the spring of 1976. Lower values on this scale represent judgments that children are less aggressive. Tests for linear trends over time were significant. Aggressiveness ratings declined for both PDC and comparison groups over time, but they declined at a faster rate for the comparison group. The presence of a linear trend is also suggested by CRS-2 curves in Figure 17d. This interpretation of the group effect, however, should be made cautiously given the significant group-by-site interaction also found.

Table 30

Results of Multivariate Analyses of Covariance with Child Outcome Measures from Fall 1976 through Spring 1979 as Repeated Measures: Sample Tested in English at Entry (Maximum n = 480)[1]

Number of Occasions	Instrument	N	Group Main Effect Over Time	Group-by-Site Interaction Over Time
Variables Measured on four Occasions	BSM-English	431		
	PIPS	424		
	POCL-1	356		
	POCL-2	356	$p<.05$	
	Verbal Fluency	424		$p<.05$
Variables Measured on three Occasions	CRS-1	326		
	CRS-2	356	$p<.05$	$p<.05$
	CRS-3	359		$p<.05$
Variables Measured on two Occasions	PIAT-Math	424		
	PIAT-Reading	398		$p<.05$

[1]Only effects reaching significance at $p<.05$ are indicated.

Turning to the interaction, within-site comparisons of linear trends indicated that significant group differences occurred only at the Michigan site where ratings of aggressiveness declined much more rapidly for the comparison group than for the PDC group from Head Start through first grade.

- Verbal Fluency interaction. This finding basically replicates the findings for questions 3 and 4, indicating that the differences observed in spring 1979 have existed over time. At the Maryland site, PDC children have consistently over time scored higher than comparison children. When scores were summed over all four occasions of measurement, the difference between group means amounted to .86 of a standard deviation. At the Washington site, the opposite trend was found: comparison children have consistently scored higher than children in the PDC group. There the difference between group mean total scores was .5 of a standard deviation. The presence of mean differences in Verbal Fluency at the Maryland and Washington sites since fall of the Head Start year suggests that these group differences may well result from pretreatment factors rather than later educational experience.

- CRS-3 interaction. CRS-3 "Dependency" scale ratings were made on three occasions beginning in spring 1976. Higher values on this scale represent higher levels of dependency. Post hoc analyses indicated that the overall interaction effect resulted from linear trends differentiating PDC and comparison groups at two sites--Connecticut and Florida. Specifically, at both of these sites ratings of dependency declined steadily and markedly for children in the comparison group, but little or not at all for the PDC group.

- PIAT-Reading interaction. The PIAT-Reading subtest was administered twice--at spring kindergarten and spring first grade. The group-by-site interaction was found to have two dimensions. First, group mean total scores--summed across two occasions--differed significantly at two sites: Michigan and Florida. In both sites, comparison children outscored PDC children on average across two occasions. Second, rates of gain from spring kindergarten to spring first grade differed significantly for PDC and comparison groups at three sites. In Colorado, PDC children gained more from kindergarten to first grade than comparison children; in California and Michigan, comparison children gained more.

Other Child Impact Analyses

In the preceding section, differences between PDC and comparison groups on child outcome means for grade 1 were explored. Those analyses of differences in central tendency are complemented in this section by analyses of differences between groups on summary score distributions, as well as by item-level analyses.

Distributional analyses of summary scores. The distributions of summary scores for each treatment group were compared for each child outcome measure. The results of these comparisons are summarized in Table 31; Figures B1 to B14 of Appendix B portray score distributions by treatment group for each measure.

Distributions of summary scores for the two treatment groups are compared by dividing the distribution for the full study sample available for each outcome measure into units of equal size on the measure, in such a way that there are about ten divisions spanning the full range of the measure. For instance, the distribution of scores for the grade 1 PIAT-Reading was divided into nine bands of size five points spanning the range between 2.5 and 47.5 points; the actual obtained score distribution ranged between five and 46 points. Once the distributions are divided in this fashion, the numbers of cases for each treatment group in each band are compared by means of a χ^2 test. If the distributions show significant overall differences, the bands at which such differences occur are identified through the use of Goodman's intervals test (Goodman, 1965).

Differences in summary score distributions between groups were found for five of fourteen grade one child outcome measures: BSM-English, PIAT-Reading, POCL-2 "Sociability," CRS-1 "Self-Assurance," and CRS-2 "Aggressiveness." Differences between distributions for all measures are associated with changes in the location of central peaks or distribution modes, rather than changes in range. Thus, differences in distributions for the BSM-English, PIAT-Reading and POCL measures are associated with central peaks at slightly higher score levels for comparison than for PDC children. For the two CRS measures, peaks occur at slightly higher levels for PDC than for comparison groups, although it must not be forgotten that for the CRS-2 "Aggressiveness" scale higher scores indicate higher levels of "aggressive" behavior. Distributional differences are relatively minor, even though statistical tests show such differences to be significant.

Analyses of item-level responses. Item-level comparisons of response frequencies for PDC and comparison children were conducted for all child outcome measures. Differences were found for items from eight measures--the two PIAT subtests, the BSM-English and BSM-Spanish, MSCA Verbal Fluency, CRS-1 "Self-Assurance," CRS-3 "Dependency," CRS-4 "Academic Motivation." Response category differences, not precisely comparable to item-level differences, were found for the PIPS.

Analyses of PIAT item-response frequencies take into account the fact that no children were administered all 84 items of either subtest, since this individually administered measure has defined test-start and test-end criteria. In the PIAT-Math, 35 items were attempted by at least ten percent of all children responding; of these, five showed differences by group in the percentages of children giving correct responses (see Table 32). Four of these differences favored the comparison group, while one favored PDC. The items included a word problem (no. 18), knowledge of the days in the week (no. 19), the ability to tell time (no. 21), a division

Table 31

Comparison of Summary Score Distributions by Treatment Group for Spring 1979 (Grade 1) Child Outcome Measures

Measure	Result of Significance Tests	Score Ranges at Which Significant Differences Occur[1]	Overall Difference
BSM-English	X^2 = 20.0, 9 d.f.; P = .019	-2.3 to -1.7: P>C -1.7 to -1.0: P>C -1.0 to -0.3: C>P	C peaks higher than P
PIAT-Reading	X^2 = 16.7, 8 d.f.; P = .033	-0.9 to -0.1: P>C	C peaks higher than P
POCL-2 "Sociability"	X^2 = 15.1, 6 d.f.; P = .020	-0.3 to +1.0: C>P	C peaks higher than P
CRS-1 "Self-Assurance"	X^2 = 16.0, 9 d.f.; P = .068 [Max. likelihood test, P = .049]	-0.9 to -0.2: C>P +0.4 to +1.1: P>C +2.4 to 2.6 P>C	P peaks higher than C
CRS-2 "Aggressiveness"	X^2 = 16.4, 8 d.f.; P = .038	+1.2 to +1.8: P>C	P peaks higher than C

Note: Only measures for which distributions differ are included in this table. In standard deviation units from the overall mean; P = PDC, C = Comparison.

[1]Distributions are illustrated in Figures B1-B14 of Appendix B.

Table 32

Item Response Frequencies by Group

PIAT-Mathematics

Item Number	Brief Description	Total Responding	% Correct: PDC	% Correct: Comp	P (Fisher's Exact Test)
18	How many pennies in a dime, nickel and penny?	492	47	55	.043
19	How many days in a week?	466	32	40	.040
21	Which clock shows 8?	386	55	70	.002
29	How many nickels = 40¢?	193	30	18	.050
34	Six boys had 5 pennies each. How many pennies altogether?	72	27	51	.009

PIAT-Reading

Number	Word Read	Total	% Correct: PDC	% Correct: Comp	P
20	Play	434	72	79	.042
21	Jump	446	74	86	.002
22	Kitten	432	42	60	.0001
23	Wagon	403	27	45	.0001
24	Fishing	402	39	54	.002
25	Brook	404	23	32	.039
32	Flour	112	27	46	.040

problem (no. 29) and a multiplication problem (no. 34). For the last two items; response levels were low and the effects of guessing could well be strong; the items certainly seem well above usual grade one performance levels. PDC children surpassed comparison children in the proportion of correct responses on item number 29.

In the PIAT-Reading, 28 items were attempted by at least ten percent of all children responding. Of these, seven showed differences by group in the percentage of children responding correctly (see also Table 32). All differences in items favored the comparison group. The items all deal with reading words, and start with the second such item of the test: "play" (no. 20), "jump" (no. 21), "kitten" (no. 22), "wagon" (no. 23), "fishing" (no. 24), "brook" (no. 25) and "flour" (no. 32).

Overall analyses identified two sites contributing to central-tendency differences between groups that could help explain these item-level differences. One site had relatively low values and no comparison group; and the other accounted for a significant site-by-treatment interaction (see Table 25 in the previous section). Item-level analyses for items showing significant differences were repeated, leaving these two sites out. Although differences between groups diminished and in fact became nonsignificant for two items, significant differences still remained for items 21, 22, 23, 24 and 32.

Eighteen responses were evaluated in the BSM, and for the English version, there were between-group differences on six items. For all six, comparison group children showed higher percentages of responses rated as correct than did PDC children; differences in the percentages ranged from five to eight percent. The test items showing differences are listed in Table 33. Children responding to the BSM-Spanish displayed between-group differences on two of the eighteen questions; for both, PDC children had higher frequencies of correct responses than did comparison children.

In the MSCA Verbal Fluency measure, the task is to name as many objects of a given type as possible in a fixed span of time. For each of four noun types (names of persons; animals; things to eat and toys), our analyses compared the distribution of children in each group by the number of objects named. The distributions were different for names of persons; more PDC than comparison group children named none or only one person.

Of eleven items composing the CRS-1 "Self-Assurance" scale, two showed significant differences between groups. The item COOPERATES AND SHARES WITH OTHERS showed PDC children higher in rated frequency than comparison group children. The item ATTEMPTS TO SOLVE SOCIAL PROBLEMS WITH LITTLE ADULT ASSISTANCE showed a difference in response distribution, with PDC children having a relatively flat distribution, while ratings for comparison children peak at central values.

Of the two items forming the CRS-3 "Dependency" scale, IS CONTROLLED OR INFLUENCED BY OTHERS showed PDC children to be displaying this behavior more frequently than comparison group children.

Table 33

Item Response Distributions by Group, BSM-English and BSM-Spanish

Item Number	Question	Total Responding	% Correct: PDC	% Correct: Comp	P (Fisher's Exact Test)
BSM-English					
6	(Point to both houses:) What are these?	546	79	87	.007
8	(Point to doors of both houses at once:) And (what are) these?	546	86	92	.014
15	(Point to eyes of both red fish:) And why do you think their eyes are open?	546	66	72	.049
16	Are the fish wet? Why?	545	67	74	.048
19a	(Point to girl:) What is the girl doing? Do you think she's happy? Why?	546	94	97	.045
22	(Point to plate:) What happened to the king's food?	544	74	81	.023
BSM-Spanish					
16	If he had not taken off his shoes (pause), what would have happened to them?	92	24	9	.044
24	What would have happened if the dog had not eaten the food?	91	38	15	.014

Of three items composing the CRS-4 "Academic Motivation" scale, the item COMPLETES ASSIGNMENTS showed PDC children displaying this behavior relatively more frequently than comparison group children.

Finally, the PIPS measure could be analyzed into twenty-five task-relevant response categories, each denoting an alternative strategy for solving an interpersonal problem. Nineteen of these categories were mentioned at least once by ten percent or more of the children responding. Of these nineteen, one showed differences in the number of children in each treatment group that mentioned the category at all (AUTHORITY REFERENCES, e.g., "I'll tell my mother and she'll come and take it from you."; more PDC than comparison children mentioned this strategy at least once). In two other categories, the numbers of children mentioning the category did not differ between groups, but the number of times the strategy was mentioned was higher for comparison than PDC children for both categories (ASK FOR IT, e.g., "Can I have it?", and TRADE A TOY FOR IT, e.g., "I'll give you this car if you give me the ____").

Summary of item-analysis findings. Only the findings for the PIAT-Reading appear to be consistent with other analyses conducted, and to convey a clear interpretation. Differences in response frequencies for some PIAT-Reading items favored the comparison over PDC group. These differences appeared only for some of the items that require children to correctly pronounce written words of increasing phonologic complexity and decreasing familiarity. The difference identified is small but appears reliable even after some site-level variation is removed.

The interpretation of item-level differences for the two BSM measures is less clear. Differences in response frequencies are small, and the items do not appear particularly difficult for the overall child population. Extensive interpretation of these differences is not justified for three measures: covariate adjustments for entry-level score differences tend to remove differences in grade one BSM outcomes, the set of items is not particularly coherent, and differences are small.

For the remainder of the instruments, also, extensive interpretation of differences in item response frequencies appears inappropriate. In the PIAT-Math, only three of the five items for which differences appear are well within the typical performance range for this ability level, and these items show little coherence in content. Differences found for the CRS scales are composed of four items, of which two favor PDC, one appears to favor the comparison group and one is difficult to interpret. Finally, response category findings on the PIPS do not support straightforward interpretations.

Overview of Impact Findings

The various findings discussed in this section on the impact of PDC on grade one child outcomes, have been consolidated. Table 34 summarizes the child outcome findings across the several approaches used to assess differences in central tendency, distributions and item-level responses for each measure. The findings appear to be replicated with fair consistency. In summary, our assessment of the impact of the PDC program on children's measures of outcome at the end of grade one is that a small number of differences between PDC and comparison groups overall can be identified, but they are generally of small magnitude and not sufficient to be educationally meaningful. The measures themselves show acceptable levels of variability between children for the task of establishing between-group differences, although most of the variation in measures is explained by within-group attributes, the principal attribute being site and site-related effects. Finally, the alternative analytic approaches used identified the same overall assessment findings. The fact that small differences can be consistently identified by various approaches provides some reassurance that, if there were educationally meaningful differences in domains measured by our instruments, we could detect them.

The Relation of Other Study Measures to Grade One Child Outcomes, Independent of Treatment

Here we consider, in a preliminary way, the relation between variables in parent and teacher domains of measurement in the PDC evaluation study and variables in the child outcome domain. Measurements in spring 1979 offer the first opportunity in this study to consider the associations between variables in different domains, since this is the first occasion in which data in the parent and teacher domains have been collected.

The analyses and findings presented here must be viewed as preliminary because formal specification of a model for relations between various PDC domains is still in an early stage of development. We simply do not know, for instance, whether it is reasonable to expect that significant changes in teacher behavior can be associated with significant changes in child outcomes measured in the same year; effects might well be lagged. Further, the ways in which associations between teacher and classroom variables and child outcome variables can be explored need further thought. Since PDC follows the development and school experiences of a sample of children, classrooms and teachers are not represented in a balanced way in the child sample. Thus, for some classrooms there may be eight children in the sample, while for others there is only one. For none of the classrooms in the study are all of the children in the classroom represented, or even half of the children.

Table 34

Summary Of Findings Contrasting PDC and Comparison Groups on Spring Grade 1 (1979) Child Outcome Measures, Using Different Analytic Procedures

	Analysis of Adjusted Data:[1]		Unadjusted Analysis of Variance	Distributional Tests	Item Analyses
	Analysis of Covariance, Each Measure	Multivariate Analysis of Covariance, all Measures			
BSM-English		P<C (.13s.d.)	P<C (.22s.d.)	C peaks slightly higher	6 of 18 items differ all differences P<C on % correct.
BSM-Spanish					2 of 18 items differ all differences P>C on % correct.
PIAT-Reading	P<C (.19s.d.)	P<C (.19s.d.)	P<C (.34s.d.)	C peaks slightly higher	7 of 28 items differ all differences P<C on % correct.
PIAT-Mathematics					5 of 35 items differ 4 differences P<C one P>C on % correct
MSCA-Verbal Fluency					More P than C children gave zero or one person names.
PIPS					
POCL-1: "Task Orientation"					
POCL-2: "Sociability"				C peaks slightly higher	
CRS-1: "Self-Assurance"				P peaks slightly higher	4 of 11 items differ two favor PDC and one is difficult to interpret.
CRS-2: "Aggressiveness"				P peaks slightly higher	
CRS-3: "Dependence"		P>C (.11s.d.)	P>C (.18s.d.)		1 of 2 items differ: PDC higher than C on rated frequency.
CRS-4: "Academic Motivation"		P>C (.19s.d.)			1 of 3 items differ: PDC higher than C on rated frequency.
CI-1: "Attitude toward Teacher"					
CI-2: "Interest in Reading"					

[1]Only for sample tested in English at entry. Data adjusted for effect of 8 covariates.

For these reasons, this section and the following one explore the relationships between variables in different domains in a highly tentative way. The purpose of these two sections is to document the fact that some associations between variables in different domains do exist, in order to suggest that further, more systematic exploration is warranted. Only a few of the possible variables have been used in the analyses; analytic approaches are deliberately kept simple, in order to permit direct interpretation to the extent possible.

After a brief discussion of the variables and analytic approaches used, we outline and discuss associations found between family characteristics and child outcomes. After that, issues in the analysis of relations between teacher-domain and child outcome variables are outlined, followed by a sketch of the associations found between teacher background variables and child outcomes, and then of associations between teacher outcomes and child outcomes.

The Association Between Family Characteristics and Child Outcomes

Variables and analytic method. Four variables defining family characteristics were examined for their relation to child outcome variables: mother's education, family annual income, family structure (as single-parent or two-parent families), and mother's employment. Thirteen child outcome variables for spring 1979 (grade 1) were used.

Variable associations were examined through the use of product-moment correlations and analysis-of-variance approaches. Child outcome variables were not corrected for initial differences, so that the full analytic sample could be used in these preliminary analyses.

Parent characteristics and child outcomes. Table 35 summarizes the relationships found between family characteristics and child outcomes, independent of treatment. The most consistent pattern of associations occurred between mother's education and the child outcome variables: twelve of the fourteen variables showed significant positive associations with the highest educational level attained. Correlations were low but statistically significant, ranging from .12 to .23. When the sample of families was divided into those with less than high school and high school or better on mother's educational level, differences between these groups were significant. The direction of differences favored families with higher levels of education. No conclusion about the relation between maternal education and child outcome above and beyond the effects usually attributed to social status can be reached at this point, however, since education is a component in usual definitions of social status.

Table 35

Relation of Family Characteristics to Grade 1 Child Outcomes, Disregarding Treatment

Child Outcome Variables	Family Characteristics			
	Mother's Education	Family Annual Income	Family Structure	Mother's Employment
BSM-English	r = .135 1<2 (P=.003)	r = .182 1<2 (P=.010)	1<2 (=.009)	
BSM-Spanish				
PIAT-Reading	r = .180 1<2 (P=.0001)	r = .219 1<2 (P=.0002)	1<2 (P=.005)	
PIAT-Mathematics	r = .217 1<2 (P=.0001)	r = .197 1<2 (P=.0002)	1<2 (P=.021)	
PIPS	r = .189 1<2 (P=.0001)			1>2 (P=.008)
POCL-1: "Task Orientation"	r = .138 1<2 (P=.0002)	r = .162 1<2 (P=.022)		1>2 (P=.047)
POCL-2: "Sociability"	r = .118 1<2 (P=.030)			
CRS-1: "Self-Assurance"	r .134 1<2 (P=.006)	r = .125 1<2 (P=.020)		1>2 (P=.0001)
POCL-2: "Aggres-siveness"				
CRS-3: "Dependence"	r = -.140 1 2 (P=.038)			
CRS-4: "Academic Motivation"	r = .163 1<2 (P=.0007)	r = .120 1<2 (P=.011)		1>2 (P=.0001)
CI-1: "Attitude toward Teacher"	r, not signifi-cant 1<2 (P=.025)			
CI-2: "Interest in Reading"	r = .228 1<2 (P=.002)	r = .144 1<2 (P=.040)		1>2 (P=.008)
Maximum n:	455	419	459	459
Dichotomized Variable Category Description	1: less than High School 2: H.S. or better	1: Less than $8000.- 2: $8000.- or more	1: single-parent family 2: two-parent family	1: mother employed 2: mother not employed

Note: Table entries are product-moment correlation coefficients and direction of differences and P levels for F-tests; only entries at X = .05 or lower levels are included.

Seven of the fourteen child outcome variables showed significant associations with annual family income, another component of social status. Once again, correlations were low, and the direction of relationships was as one might expect: higher income levels are associated with more favorable outcomes.

Significant associations between the family structure variable and child outcomes occurred only for the three outcome measures most clearly related to traditional views of academic progress: the BSM-English and the two PIAT subtests. Children from single-parent families did less well than children from two-parent families on these three variables. Interestingly enough, however, no differences between these two groups appeared on any of the other outcomes related to other aspects of social competence, such as interpersonal style (the PIPS, CRS-2 and CRS-3, CI-1), task orientation (POCL-1), or academic motivation (CRS-4, CI-2).

Finally, significant associations occurred for five child outcome variables with mother's employment. Children in families in which mothers are employed scored more favorably than children of families in which mothers are not employed. The pattern of outcome variables showing differences is somewhat surprising in its consistency, and almost opposite to that shown for family structure: no differences appear in the more traditionally academic measures, but clear differences on variables related to interpersonal style (PIPS), task orientation (POCL-1, CRS-1), or academic motivation (CI-2, CRS-4). The possibility that mother's employment is serving as an additional indicator of social status in this sample should not be ignored in considering interpretations of this finding.

The Association Between Teacher Background Characteristics, or Teacher Outcomes, and Child Outcomes

Issues in cross-domain analyses. The present report presents initial explorations of relations between teacher and classroom domain measures and child outcomes. However, a number of issues obscure the selection of the appropriate units of analysis in crossing from the teacher or classroom domain to the domain of child measures.

By way of contrast, analyses exploring the association between family characteristics and child outcomes offer no ambiguity in definition of the appropriate unit for analysis: for each child in the sample there is one family, and vice versa. For each teacher or classroom, on the other hand, there is generally more than one child. Further, the number of children for each teacher or classroom varies widely (and varies along different ranges for PDC and comparison groups). This situation poses a number of issues for which resolution is not straightforward. The issues can be summarized in the following question: to what extent is it reasonable to expect that associations between teacher characteristics or outcomes and child measures can be reliably identified when samples of children for each teacher are small, of varied sizes and possibly unbalanced?

In spite of the potential problems, it does seem reasonable to assume that the children in the PDC evaluation study sample have been allocated to their teachers in nonsystematic ways. In the light of this assumption, it would seem worthwhile, as a first approximation, to explore whether there are any associations across domains. If significant associations are found, further analyses using strategies to (1) balance samples of children by groups or (2) select subsamples of classrooms with the most children or with similar distributions of children per classroom might be expected to show even stronger associations. In other words, assessment of associations between the teacher and child variables without efforts to restructure the sample provide highly conservative tests of association between these variable domains--if the assumption of nonsystematic assignment of children to teachers is warranted. Since the validity of this assumption has not yet been tested, associations across these domains, are highly tentative at this point in the analytic process.

One further point deserves mention. Two alternative approaches are used to examine associations between teacher and child domains; using teachers and using children as the unit of analysis. The child level uses repeated non-independent data (the teacher domain data) and thus overestimates effects by using more degrees of freedom than are really available. The teacher or class level weights each child's score differently, depending on the number of children in a given classroom. There are clear problems with both approaches. For the exploratory purposes of this report, both approaches have been used and findings are presented only when both approaches agree on the direction and significance of associations.

Variables and analytic methods. Three teacher background characteristics and nine teacher outcome variables were examined for their association with child outcome variables. The teacher variables are listed below. Definitions for these variables are offered in Volume I of the present report. Thirteen child outcome variables for grade 1 (spring 1979) were examined.

Teacher background characteristics

- number of years teaching at this school
- educational level
- teacher gender

Teacher outcomes

- frequency of committee or task force participation
- extent of parents' educational activities in the classroom
- attitude toward increased parent involvement in the classroom
- frequency of use of community resources

- extent of training in child development
- Factor 1: program adaptation of individual children
- Factor 2: extent of structuring of activities in language arts and mathematics
- Factor 3: teacher's efforts to involve parents and the home in class activities
- Factor 4: extent of individualization of activities in language arts and mathematics

Variable associations were examined through the use of product-moment correlations, where appropriate, and analyses of variance. Child outcome variables were not corrected for entry-level differences, so that all of the available analytic sample could be used in these exploratory analyses. For each variable association examined, separate analyses were conducted, using teachers and children as the analytic unit. We present the findings below only for those associations where the results for both levels of analysis (child and teacher) were in agreement as to significance and direction. Among the 169 associations examined, there were 10 cases where the results were inconsistent: in nine of the cases the findings were significant at the child level but did not reach significance at the teacher level of analysis.

Teacher background characteristics and child outcomes. Table 36 summarizes the associations found between the three teacher background variables and grade 1 child outcome measures.

The variable number of years teaching at this school appears to be negatively associated with teacher ratings of child aggressiveness and dependence (CRS-2 and CRS-3 scales). Teachers with fewer years teaching showed a tendency to rate children higher on these two scales. This finding might be interpreted as indicating a degree of teacher bias in the assessment of children's interactive characteristics in the classroom--a bias whose direction changes with experience. No other child outcome variables based on teacher ratings showed consistent relations with teacher background characteristics.

Teacher's educational level showed a significant association with both of the scales of the POCL: the POCL-1 "Task Orientation" scale and the POCL-2 "Sociability" scale. In both cases, teachers with the highest educational levels tended to have children in their classes rated higher on these scales by the testers. The same pattern of association appeared for the CI-2 "Interest in Reading" scale: teachers with higher educational levels were linked to children with higher levels of interest in reading.

Finally, teacher's gender appeared related to levels of interest in reading: male teachers were linked to children with less interest in reading than were female teachers. This does not seem to be a site-specific effect, since the male teachers were spread out over six of the eleven sites. However, the small number of male teachers in the sample (eleven) adds to the difficulty of interpretation of this finding.

The child outcome variables which show no relationships with teacher background may be of as much interest as the child outcome variables which do show associations with teacher background. None of the more traditionally "academic" child outcome variables (the BSM-English and Spanish and the two PIATs) show consistent associations with the three teacher background variables examined. Instead, associations are found between teacher background characteristics and task-related and interpersonal competence assessments.

Of thirty-nine possible associations tested, six show relations that are statistically significant at the $\alpha = .05$ level or better--much better results than those one might expect at random. Table 36 presents these results. These findings are encouraging; since they are based on two separate (but not independent) analyses in each case, and since there is reason to believe that the degree of association found is a highly conservative estimate of the relation that might actually exist. It appears worthwhile to continue to examine the possible relationships between teacher background characteristics and child outcomes.

Teacher outcomes and child outcomes. Table 37 summarizes the associations found between teacher variables that might be considered modifiable by the PDC program and child outcomes. As the table shows, they are few and somewhat scattered.

Interviewers' ratings of the extent to which teachers permitted parents to undertake educational activities in the classroom (as opposed to inviting them only for routine parent conferences) was positively associated with scores on the BSM-Spanish, in spite of the fact that the sample of children who took the BSM in Spanish (and of their teachers) was quite small: 82 children, and 22 teachers. This measure of teachers' behaviors with regard to parent involvement was not consistently associated with any other child outcome.

The rated extent of teacher structuring of activities in language arts and mathematics was associated with student scores on the BSM-English. Teachers who were rated by the observers as providing relatively more child-centered activities or as structuring a heterogeneous mix of activities for these areas had children with higher BSM-English scores than did teachers with a more teacher-centered, homogeneous approach.

A second composite measure derived from the Teacher Interview global ratings, the extent of teachers' efforts to adapt the educational program to the needs of individual children, showed significant associations with

Table 36

Relation of Teacher Background Characteristics to Grade 1 Child Outcomes, Disregarding Treatment

Child Outcome Variables	Teacher Background Characteristics		
	Number of Years Teaching at Their School	Educational Level	Teacher Gender
BSM-English			
BSM-Spanish			
PIAT-Reading			
PIAT-Mathematics			
PIPS			
POCL-1: "Task Orientation"		T: 1,2<3 (P=.029) C: 1,2<3 (P=.010)	
POCL-2: "Sociability"		T: 1<2,3 (P=.040) C: 1<2,3 (P=.017)	
CRS-1: "Self-Assurance"			
CRS-2: "Aggressiveness"	T: 1>2 (P=.002) C: 1>2 (P=.0004)		
CRS-3: "Dependence"	T: 1>2 (P=.010) C: 1>2 (P=.0001)		
CRS-4: "Academic Motivation"			
CI-1: "Attitude toward Teacher"			
CI-2 "Interest in Reading"		T: 1,2<3 (P=.020) C: 1,2<3 (P=.011)	T: 1<2 (P=.010) C: 1<2 (P=.0004)
Maximum n:	159 Teacher-level 481 Child-level	161 Teacher-level 497 Child-level	161 Teacher-level 496 Child-level
Variable Category Descriptions:	1: five or fewer years 2: six or more years	1: 4-year college degree or less 2: credit toward M.A. 3: M.A. or better	1: male 2: female

Note: Table entries are the direction of differences and P levels for t or F tests; only entries at X = .05 or lower levels on both analyses using teachers and children as units of analysis are included. T = teacher-level, C = child-level.

Table 37

Child Outcomes, Disregarding Treatment

Child Outcome Variables	Teacher Outcome Variables		
	Parents' Educational Activities in the Classroom	Factor 1: Program Adaptation to Individual Children	Factor 2: Structuring of Language Arts and Math Activities
BSM-English			T: 1,3,2 (P=.0001) C: 1,3,2 (P=.0001)
BSM-Spanish	T: 1,2,3 (P=.036) C: 1<2<3 (P=.0004)		
PIAT-Reading			
PIAT-Mathematics			
MSCA-Verbal Fluency			
PIPS			
POCL-1: "Task Orientation"			
POCL-2: "Sociability"			
CRS-1: "Self-Assurance"		T: 1,3,2 (P=.005) C: 1<3,2 (P=.003)	
CRS-2: "Aggressiveness"			
CRS-3: "Dependence"			
CRS-4: "Academic Motivation"		T: 1,3,2 (P=.049) C: 1<2,3 (P=.016)	
CI-1: "Attitude toward Teacher"			
CI-2: "Interest in Reading"			
Maximum n:	159 Teacher-level 489 Child-level	162 Teacher-level 497 Child-level	162 Teacher-level 497 Child-level
Variable Category Descriptions:	1: none of the parents 2: some 3: most of the parents	1: low individualization 3: high individualization	1: teacher-centered, homog. 3: child-centered, heterog.

Note: Table entries are the direction of differences and P levels for F tests; only entries for which $\alpha = .05$ or lower levels on both analyses using teacher and children as analytic units are included. T = teacher-level, C = child-level.

two ratings of the children by teachers: the CRS-1 "Self-Assurance" scale and the CRS-4 "Academic Motivation" scale. These associations occur between two different assessments obtained from the same source (the teacher). The implication is that those teachers who saw themselves as trying the most to adapt their educational efforts to individual needs were also the ones who rated their children highest on personal and academic competencies.

Six other teacher outcome variables showed no consistent association with child outcomes. These teacher outcome variables are : frequency of committee or task force participation; attitude toward increased parent involvement in the classroom; frequency of use of community resources; efforts to involve parents and the home in class activities; individualization in language arts and mathematics activities; and amount of training in child development.

Only four associations were consistently identified, as Table 37 shows, out of 117 possible associations. Although this number of associations could be found by chance, it is important to remember that we are using quite conservative standards for these exploratory analyses, and that the associations identified are restricted to only three teacher outcome variables. The number becomes more encouraging when we consider that four associations are significant for 39 tests on three teacher outcome variables.

The Interaction of Other Study Measures with Treatment in Relation to Grade 1 Outcomes for Children

In this section, we address briefly the following question: to what extent do variables other than treatment help explain treatment-related differences in grade 1 child outcomes? This question breaks down into two separate questions:

- Are there interactions (that is, non-additive effects) between educational treatment and parent background, teacher background or teacher outcome variables?
- Do these other measures help account for treatment-related differences in child outcome measures, or are their contributions to child outcome measures separate from the effects of treatment?

In reality, these two questions are not separate. If there are no interactions between educational treatment and the other study measures, then it is likely that the contributions of educational treatment and the other measures to explaining variation in child outcomes will be separate. The second question is included because the possibility also exists that, after other study measures have accounted for variation in child outcomes,

there will be no remaining effect of treatment. If this is the case, we will have to look for relationships from educational treatment to other study measures and then to child outcomes.

As noted earlier, the analyses conducted to explore associations between educational treatment, other study measures in the family and teacher domains, and child outcomes are preliminary. The process of construction of an analytic model is still in its early stages; only cross-sectional data across domains are available; and the sample of classrooms and teachers is not balanced with regard to the sample of families and children. The purpose of these analyses is to document the extent to which interactions occur and to identify analytic problems, to form a basis for further work.

Analytic methods. To examine the first question, the presence of interactions between educational treatment and other study measures in relation to child outcomes, we dichotomized the predictor variables, created dummy variables to represent them, and created dummy variables to represent the interaction between treatment and each other predictor. We then contrasted restricted and unrestricted regression models, in which predicted scores on each child outcome measure are used as dependent variables.

The second question was examined by constructing a multiple regression model for each of the child outcome measures. Other study measures were entered first in these regression equations, and then the additional contribution of educational treatment was examined.

Variables used. Four sets of variables were used in the analyses for this section: child outcome measures, family background characteristics, teacher background characteristics and teacher outcomes. Only those variable pairs for which significant univariate relationships with child outcomes were identified in previous sections were used in analyses. While interactions might be found where there are no main effects, both the anticipated difficulties in interpretation and the exploratory, preliminary nature of these analyses lead us to decide to restructure these analyses to an examination of interactions only where one or both of the main effects (either treatment or the other study measures) have been previously identified. The child outcome variables used include the thirteen child outcome measures used in the previous section:

- BSM-English and BSM-Spanish
- PIPS
- PIAT-Mathematics and PIAT-Reading
- POCL-1 and POCL-2 scales
- CI-1 and CI-2 scales
- CRS-1, CRS-2, CRS-3 and CRS-4 scales

Four family background characteristics include:

- mother's educational level
- annual family income
- family structure (as single-parent or two-parent families)
- mother's employment

Teacher background characteristics include:

- number of years teaching at the school
- teacher's educational level
- teacher gender

And the teacher outcome variables are:

- extent to which parents are allowed to undertake educational activities in the classroom
- program adaptation to individual children
- structuring of activities in language arts and mathematics

The Interaction of Educational Treatment and Other Study Measures with Child Outcomes

A total of sixty-three variable pairs were examined for interaction with treatment. Three interactions were identified. Since this is the number of interactions that would have been identified by chance if there were only random associations (using $\alpha = .05$), and since the interactions identified are not readily interpretable, they are reported only in summary manner. An interaction occurs between: educational treatment and mother's educational level in the PIAT-Reading, but disappears if the Georgia site (which has low scores and no comparison group, and can reasonably be interpreted as a site-specific effect) is removed; educational treatment and mother's employment with relation to the CI-2 "Interest in Reading" scale; and educational treatment and structuring of activities in language arts and mathematics in relation to the CRS-3 "Dependence" scale.

It seems reasonable to conclude that the variables examined do not have significant interactions with educational treatment in accounting for variation in child outcomes, because of the few and widely scattered interactions obtained for all those examined.

Additional Contributions of Educational Treatment to Child Outcomes, Over and Above the Contributions of Other Study Measures

In light of the exploratory nature and the complexity of these analyses, sequential multiple regressions were conducted only for the PIAT-Reading and BSM-English measures. These child outcome variables were approximately of

interval scale, had shown significant treatment-related impacts, and were of a more "academic" nature. We conducted these exploratory analyses to see whether, once variance in the outcome measures associated with parent and teacher variables and site effects were removed, effects of educational treatment persisted.

Family background characteristics were the first group of variables used as predictors. As expected, they explained significant proportions of outcome measure variance. Dummy variables for site effects then explained additional amounts of variance. Teacher outcome variables were then added, followed by educational treatment. Variables from each of these groups made significant additional contributions. An example of such a regression equation sequence is shown in Table 38, which illustrates that both teacher outcomes and educational treatment make separate additional contributions to explaining variation in the grade one PIAT-Reading, after the effects of site and family background are removed. If this finding is confirmed, it would suggest that these particular changes in teacher characteristics are not the ones contributing to educational treatment differences. However, such conclusions are highly speculative at this point: conclusions are unwarranted until a larger set of teacher and classroom-level outcome variables are available, until their relation to educational treatments in both PDC and comparison schools is understood, and until issues in selection of appropriate levels of analyses for child outcomes are confronted.

Table 38

Results of Stepwise Multiple Regressions on
Grade 1 PIAT-Reading Scores (n=329)

Variable Sets	R^2 After Inclusion	Additional F	Contribution P
1. Parent Background (2 var.): Mother's Education and Family Income	.06	11.03	.0001
2. Site (3 dummy variables)	.10	3.03	.026
3. Teacher-level Outcomes (4 dummy variables)	.14	3.65	.007
4. Educational Treatment (1 variable)	.17	11.10	.001

Notes: Data for Georgia site are excluded.
The order of variable sets in the table reflects the order of inclusion in successive regression equations, in which variable sets appearing earlier are fixed.

VI

SUMMARY: CHILD OUTCOMES AT FIRST GRADE

The organization of this chapter parallels that of Chapter V. PDC's impact on child outcome measures is examined first; succeeding sections examine the study's finding on the relation of measures across domains independent of educational treatment, and the extent to which other variables modify the relations between educational treatment and child outcomes. A final statement summarizes the findings in general terms and analyzes their implications for future data collection and analysis.

PDC's Impact on Child Outcome Measures

The analytic approach to child o tcome measures involved a sequence of three questions, each of which was aimed at a different type of overall comparison of test scores for children in the PDC program with children in the comparison groups. The questions were:

Question 3: *Considering outcome variables one at a time, are PDC and comparison children different from one another at the end of first grade (spring 1979)? In particular, are there group main effects or group-by-site interactions?*

Question 4: *Considering all outcome variables simultaneously, are PDC and comparison children different at the end of first grade (spring 1979)? In particular, is there a group main effect or group-by-site interactions?*

Question 5: *Considering all occasions of measurement of each child outcome variable from fall 1979 through spring 1979, are there trends and patterns over time in group differences or group-by-site interactions?*

Analytic methods used for each of these questions are discussed in detail in Chapter V. Findings are summarized below for each instrument. The summarization is restricted to analyses for the sample tested in English. The sample of children tested at some times in Spanish was analyzed separately, but complete analyses cannot be reported due to problems in interpretation associated with reduced sample size and gross imbalances in distribution across sites.

Summary of Findings for Child Outcome Measures

Peabody Individual Achievement Test-Reading. A significant main effect favoring the comparison group was found in analyses of grade 1 scores for the several outcome variables separately (question 3) and for the outcome variables considered together (question 4). The magnitude of the effect was approximately .19 of a standard deviation. When the low scoring PDC sample from the Georgia site, which had no comparison group, was removed from the analytic design for the separate analyses for each outcome variable (question 3), the overall difference between groups disappeared. No group effect was found in the analyses of trends in outcome variables over time (question 5).

Significant group-by-site interactions were found in analyses for all three questions relating to child outcomes. In each case, post hoc tests indicated that the Michigan site's comparison group had a significantly higher mean PIAT-Reading score than the PDC group. Post hoc analyses of the interaction of trends in outcome variables over time (question 5) further indicated:

- a main effect over two occasions favoring the comparison group in Florida,
- higher rates of gain from spring kindergarten to spring first grade by comparison groups in both Michigan and California, and
- a higher rate of gain from kindergarten to first grade by the PDC group in Colorado.

McCarthy Scales and Children's Abilities-Verbal Fluency. Significant group-by-site interactions were found in analyses for all three questions about child outcomes. Post hoc analyses revealed that these interaction effects were due to group differences at two sites: the PDC group was higher than the comparison group in Maryland, while the comparison group was higher in Washington.

Bilingual Syntax Measure-English. A significant effect favoring the comparison group was found in analyses (question 4) with all outcome variables considered together. The magnitude of the effect was small--.13 of a standard deviation.

Child Rating Scale-2 "Agressiveness" scale. A significant effect over time favoring the comparison group was found in question 5 analyses, trends in outcome variables over time. Post hoc analyses of a significant group-by-site interaction for this measure revealed a significant site-level difference only in Michigan. The magnitude of the effect was small--.15 of a standard deviation.

Child Rating Scale-3 "Dependence" scale. A significant effect favoring the comparison group was found in question 4 analyses, with all outcome variables considered together. The magnitude of the effect was small--.11 of a standard deviation. A significant group-by-site interaction was found in question 5 analyses trends in outcome variables over time, revealing a trend over time that favored the comparison groups in Florida and Connecticut.

Child Rating Scale-4 "Academic Motivation" scale. A significant effect favoring the PDC group was found in question 4 analyses, with all outcome variables considered together. The magnitude of the effect was small--.19 of a standard deviation.

Pupil Observation Checklist-2 "Sociability" scale. A significant effect over time favoring the comparison group was found in question 5 analyses, trends in outcome variables over time. Differences between groups did not reach statistical significance at any single point in time. The magnitude of the difference was relatively small--.25 of a standard deviation.

Spring 1979 outcome profile. Question 4 analyses, which considered all spring 1979 measures simultaneously, indicated that outcome profiles for the two groups were significantly different. Differences for particular measures, however, were inconsistent in their direction and so small (cf. the estimated magnitudes of effects described above) as to be of dubious educational consequence.

Interpretation of Child Impact Findings

Our interpretation of the findings for child outcome measures through the cohort's first grade (spring 1979) is as follows. First, the set of measures appears overall to have met the psychometric standards established for the child outcome battery; it also appears to be identifying variation among children. The bulk of such variation, however, appears associated with differences between sites and within groups, rather than being associated with differences between PDC and comparison treatment groups. Second, the analytic approaches used in this report seem appropriate for the task of identifying differences between PDC and comparison groups, since they provide fair consistency in their findings across a variety of approaches. Third, the analyses carried out discriminate between treatment groups and identify outcome differences, but these differences as of the end of first grade are too small to be educationally meaningful. On the basis of these findings, it is our sense that if there were educationally meaningful differences between PDC and comparison children within the scope of the outcomes measured, our analytic procedures would be adequate to identify them.

On the basis of our analyses of impact on child outcome measures for grade one we conclude, therefore, that there are no significant, educationally interpretable differences overall between PDC and comparison groups of children.

Some differences at the level of individual sites do appear between PDC and comparison groups. Site-level analyses and findings, however, are not appropriately conducted with the same techniques used for overall, across-site child outcome measure analyses. Covariate adjustments that appropriately correct for initial overall entry-level differences may leave large entry-level differences between groups at particular sites, and these residual differences may in turn be affecting site-level outcome differences between treatment groups. As data analyses at grades 2 and 3 are conducted we intend to consider site-level phenomena in greater depth, utilizing procedures specifically tailored to site-specific analyses.

The issue of site-specific effects and their connection to entry-level differences is related, in its implications, to another finding in this volume: that the samples of PDC and comparison children available for testing in spring 1979 are no longer fully comparable. PDC children were found to have lower entry-level test scores than comparison children. Univariate analyses revealed significant differences on two entry-level tests; multivariate analyses indicated that entry-level test score profiles were different for the two groups. Though these differences were not large, they consistently favored the comparison group. No group differences appeared on demographic variables. The issue of the changing comparability of PDC and comparison members of the study cohort will be a specific focus of analytic concern in the future; our aim will be primarily to come to understand how it happened.

It is the lack of comparability of PDC and comparison grade one groups of children on entry-level test data that made covariance-based adjustments necessary for between-group comparisons on grade one child outcomes. Although these techniques seem appropriate for cross-site analyses, they increase the complexity of interpretation of site-specific effects. The approach of conducting separate site-specific adjustments for each site (which amounts to treating each site as a separate replication of the PDC experience) has its own drawbacks. The advisability of such an approach will be examined in reference to future analyses.

The Relation of Other Study Measures to Grade 1 Outcomes, Regardless of Treatment

Preliminary analyses were conducted to assess the presence and extent of relations between child outcome measures and variables measuring aspects of family background, teacher background and teacher attitudes and behaviors potentially interpretable as program "outcomes." These analyses are viewed

as preliminary because they are based on only one year's data, because they precede the existence of a specified model of hypothesized relations between these domains, and because they make certain conservative assumptions about the appropriate units for cross-domain analyses. The findings from these analyses document the fact that cross-domain relationships can be identified for the study data base and that more systematic consideration of these relationships would be worthwhile.

Summary of Findings

Family background and child outcomes. Family characteristics show consistent and positive associations with grade one child outcome measures. The broadest range of such associations occurs for mother's education and family income. Higher levels of mother's education and family income are associated with higher scores on several of the child outcome measures. The correlations were low but consistent and significant. These findings replicate the usually identified relationship between social status and achievement, since mother's education and family income are components in the usual definitions of social status. The family structure variable (which distinguishes between single-parent and two-parent families) shows children of two-parent families scoring higher than children of single-parent families on the more traditionally "academic" child outcome measures such as the PIAT-Math, PIAT-Reading and BSM-English. The variable mother's employment shows children of families in which the mother is employed scoring higher than children in families in which the mother is not employed, on child outcome measures dealing with areas of interpersonal and task-related competency such as the PIPS, POCL-1 "Task Orientation," CI-2 "Interest in Reading," CRS-1 "Self-Assurance," and CRS-4 "Academic Motivation."

Teacher background and child outcomes. Perhaps the most notable finding of association between these two study domains is a negative one: none of the more traditionally "academic" of the child outcome measures (such as the two BSM tests or the two PIAT subtests) show any relationship to teacher background variables. Instead, measures of teacher background show relationships with child outcomes related to task and interpersonal competencies. Thus, the variable number of years of teaching at this school is related negatively to CRS-2 "Aggressiveness" (with the teachers with the most experience rating the children as least aggressive), and in the same way to CRS-3 "Dependence." The teacher's educational level is related positively to ratings of students on the POCL-1 "Task Orientation" and the POCL-2 "Sociability," and to the CI-2 "Interest in Reading."

Teacher outcomes and child outcomes. The observer's rating of the teacher's extent of structuring in language arts in mathematics activities in the classroom is associated with student scores on the BSM-English. Teachers with children with higher scores on this measure tended to be those who were rated as providing relatively more child-centered activities or a more heterogeneous activity mix, rather than those teachers rated as providing a more teacher-centered, homogeneous approach.

A second composite measure derived from the Teacher Interview global ratings, the extent of teachers' efforts to adapt the educational program to the needs of individual children, showed significant associations with two ratings of the children by teachers: the CRS-1 "Self Assurance" scale and the CRS-4 "Academic Motivation" scale. Since these relationships occur between two different assessments obtained from the same source (the teacher), the implication is that those teachers who saw themselves as trying the hardest to adapt their educational efforts to individual child needs were also the ones who rated their children highest on these competencies.

Interpretation of Variable Associations Independent of Treatment

Associations between family background variables and child outcomes are not new or surprising; perhaps their most useful purpose will be for the isolation of social status-related components of child outcome measure variation. The fact that different aspects of child outcome are tapped by the two measures of family structure and mother's employment suggests the importance of an inspection of the interaction of these two variables with each other.

For the assessments of associations between teacher-domain and child outcome variables, perhaps the most significant statement that can be made at this point in the analysis sequence is that some relationships were found. Teacher background and outcome variables seem to be associated mostly with the less traditionally "academic" outcome measures, and in particular to bear some relation to teachers' ratings of their children.

The Interaction of Educational Treatment and Other Study Measures with Child Outcomes

The research question that sparked the analyses undertaken for this section was: to what extent do other study variables modify the relation between educational treatment and child outcomes at grade one?" Our highly preliminary answer to this question is "not much." While there are at best weak effects of treatment on the child outcome variables at the first grade level, there do not appear to be interactions between educational treatment and variables in the parent and teacher domains, in relation to child outcome measures. Additional exploratory analyses were carried out of the contributions of educational treatment to the explanation of variation in child outcomes over and above the contributions of family background variables, between-site differences and teacher outcomes. These analyses seem to indicate that the effects of the educational treatment are not confounded with variables used from these domains of family background between-site differences, and teacher outcomes. If pathways between educational treatment and child outcome are to be sought, they must be sought elsewhere. Of course, many available pathways have not been explored: variables from different points in time; other variables from

the teacher and parent domains (including parent "outcomes" not explored in current analyses); and a number of variables representing data collected on the classrooms through direct observation. Most significantly, a model of the sequence of impact of PDC, sufficiently detailed to allow for specification of analyses, still remains to be fully developed. The few findings reported in these preliminary analyses indicate that such a model is required if the large variety of findings from the many domains that form the PDC study are to be formed into a coherent whole.

At the conclusion of the grade one year of schooling for the evaluation cohort, the evaluation findings for child outcomes are mixed. There are no educationally meaningful and consistent differences between PDC and comparison students on the child outcome measures. There are some differences between family background, teacher background, and teacher outcome variables and some child outcome measures. The impact of treatment (PDC versus comparison) over and above other variables in explaining differences in child outcomes is not clear. However, the number of potential relationships among study variables in any one year of schooling of the evaluation cohort, the additional complexity of the longitudinal analyses, the differential effects on PDC and comparison samples of loss of students to the longitudinal evaluation, and the technical analysis problems (such as unit of analysis questions) indicate the need for careful consideration of the possible evaluation findings, in the form of the development of one or more models of possible relationships between PDC impacts. The model described in Chapter II is the first step in that development. As details of the model are identified, analyses will be specified and conducted with the longitudinal data set from the evaluation of Project Developmental Continuity.

References

Finn, J. D. A general model for multivariate analysis. New York: Holt, Rinehart and Winston, 1974.

Goodman, L. A. On simultaneous confidence intervals for multinominal proportions. Technometrices, 1965, 7, 247-254.

Granville, A. C., Love, J. M., & Morris, M. A process evaluation of Project Developmental Continuity, Interim Report VII, Volume 3: Assessment of program impact through the Head Start year. Ypsilanti, Mich.: High/Scope Educational Research Foundation, August 1977.

Granville, A. C., McNeil, J., Shelly, M., Morris, M., Meece, J., & Wacker, S. A process evaluation of Project Developmental Continuity, Interim Report VI: Recommendations for continuing the impact study. Ypsilanti, Mich.: High/Scope Educational Research Foundation, March 1977.

Granville, A. C., Morris, M., Love, J. M., & Hepler, N. Evaluation of Project Developmental Continuity, Interim Report IX: Assessment of program impact through the kindergarten year. Ypsilanti, Mich.: High/Scope Educational Research Foundation, October 1979.

High/Scope Educational Research Foundation. Project Developmental Continuity: Field procedures manual. Ypsilanti, Mich.: Author, March 1979.

APPENDIX A

Child Development Measures

I. CHILD MEASURES

Purpose of Measures

During the course of the PDC evaluation a variety of child measures have been selected and/or developed to assess program effects in the key domains of children's behavior and development. The original battery of measures was identified during 1974-75, pilot-tested during 1975-76 and used at the Head Start level in 1976-77. With a few additions and deletions, this set of measures has been used in each testing period from Fall 1976 through Spring 1978. In Spring 1979, however, several changes were made to accommodate the increasing age of the children and to be responsive to the expectations of public schools. Before describing these changes and the rationale for them, this section summarizes the purposes underlying the original battery (a more detailed description of the rationale for the original battery may be found in Interim Report II, Part B, 1975).

The Original Battery

The measures administered in Fall 1976 were selected according to a number of criteria. These included (1) practical considerations (e.g., available for use in Fall 1975, appropriate for administration by trained paraprofessionals, reasonable testing time); (2) psychometric qualities (e.g., validity, stability and internal consistency, representativeness of the standardization sample); (3) relevance to PDC (appropriate age range, Spanish adaptation available, relevant to program goals, likely to demonstrate program effects); and (4) past use (i.e., used in previous national evaluations or large-scale studies). In addition, one of the major goals of the evaluation during 1974-75 was to be sensitive to children's "social competence." Consequently, several sources were reviewed for guidance on defining and measuring social competence. These included the Office of Child Development's 1972 statement (i.e., "an individual's everyday effectiveness in dealing with his environment..."); the discussion of 29 competencies by Anderson and Messick (1974); reviews of the Rand Corporation for its design of a national evaluation of Head Start (Raizen & Bobrow, 1974); and research on social-emotional behavior by White and Watts (1973) and Ogilvie and Shapiro (1973). An attempt was also made to take PDC program goals into consideration by polling PDC staff regarding their expectations for children.

Over a two year period the initial battery of 17 tests, two observation instruments and two rating scales (plus measures of height and weight) were modified so that by spring of the test cohort children's Head Start year

the battery consisted of five tests of cognitive or language development, four measures of social-emotional development (one test, two rating scales and a child observation system), one psychomotor measure and measures of height and weight. Changes in the battery over time were based on assessments of six criteria at each testing point: internal consistency reliability, validity, sensitivity to change, relevance to "social competence," suitability for use in the higher grades and ease of administration.

The Spring 1979 Battery

Although the original measures were selected with some attention to their continued suitability as children became older, the main emphasis was on their use in Head Start. One of the chief concerns in program evaluations at the elementary school level is school achievement, however, which, of course, was not represented in the PDC Head Start level battery. Thus, for Spring 1978 testing, when the test cohort was in kindergarten, an achievement test was added to the original measures. The rest of the battery remained similar to the Head Start battery. At the first grade level (Spring 1979) several additional modifications of the battery were made. These included adding parent ratings of children's progress in school and attitudes toward school; adding a child interview to tap children's attitudes toward their teachers and toward reading; adding an additional scale to the Child Rating Scale to tap academic motivation; recording attendance, retention in grade, special education placement, and extent of mainstreaming; and deleting the child observation system, arm coordination, verbal memory and draw-a-child. The rationale for each of these changes is presented here, following a discussion of the basic framework adopted for the Spring 1979 battery.

During the first phase of the PDC evaluation, child measures were classified as social-emotional, psychomotor, and cognitive-language. In preparation for the current phase, considerable discussion took place between High/Scope and ACYF staff regarding the domains that should be represented in the child battery. These discussions were necessary because, although there has been considerable research and evaluation at the Head Start level, there are few precedents for studies that follow Head Start children into the elementary grades. The classifications of measures used in three evaluations were reviewed (see Figure 1) and found to represent very different patterns of organization. But in discussions of domains several themes continually emerged. First, in a program concerned with enhancing children's effectiveness in school, some measure of school achievement must be included; regardless of the shortcomings of typical standardized achievement tests, these tests represent politically important criteria for programs such as PDC. Second, ACYF and High/Scope are interested in going beyond the politically important criteria and selecting measures that are important from a program perspective. Thus, in the

Domains and Measures Included in Head Start and Follow Through Evaluations Where Elementary School Children Were Tested

Follow Through Evaluation (Stebbins, et al., 1977)	Home Start Follow-up Study (Abt & High/Scope)	Head Start Transition Study (Royster, et al., 1978)
Basic Skills	**School Achievement**	**Academic Achievement**
Word Knowledge (MAT)	Mathematics (PIAT)	Reading (WRAT)
Spelling (MAT)	Reading Recognition (PIAT)	Spelling (WRAT)
Language (MAT)		Numbers (WRAT)
Math Computation (MAT)	**Social-Emotional**	Academic potential (Schaefer)
Cognitive-Conceptual	Self-concept (Stephens-Delys)	**Prosocial Adaptive Behaviors**
Reading (MAT)	Attitude toward peers (PSAS)	Sociability (Values Inventory for Children)
Math Concepts (MAT)	Attitude toward school (PSAS)	Conformity (VIC)
Math Problem-Solving (MAT)	Attitude toward family (PSAS)	Friends named (Child Interview)
Nonverbal Problem-Solving (Ravens)	Social problem-solving (PIPS)	Extroversion (Schaefer)
Affective-Cognitive	Task orientation (POCL)	Popularity (Schaefer)
Self-Esteem (Coopersmith)	Sociability (POCL)	Sociability (POCL)
Locus of Control (Intellectual Achievement Responsibility Scale [+,-])		**Unsocial, Regressive Behaviors**
		Asocial (VIC)
		Me First (VIC)
		Aggression (Beller Rating Scales)
		Hostility (Schaefer)
		Motivational Orientation
		Academic motivation (VIC)
		Absence from school (records)
		Task orientation (Schaefer)
		Autonomous achievement striving (Beller)
		Test orientation (POCL)
		Dependency Behaviors
		Closeness to adults (VIC)
		Child dependency (Beller)
		Adult dependency (Beller)
		Dependency conflict (Beller)

domain we call "academic skills and abilities" measures of cognitive and language development are included. Also because of program emphasis, a range of measures in the domain that's commonly referred to as social-emotional or affective are included. For the purpose of PDC, we divided this area into social-emotional development and learning attitudes because of increasing awareness of the importance of attitudes and motivation as concomitants of achievement and general social competence. In recognition of the importance of the growing child's physical well-being, some measures of health and nutrition are included. Finally, the classroom observation system (see Section IV of this appendix) afforded us the opportunity to obtain indices of children's classroom behavior. Figure 2 lists the measures used at first grade, organized by the four domains discussed above:

(1) academic skills and abilities;
(2) social-emotional development and learning attitudes;
(3) health and nutrition; and
(4) classroom behavior.

Descriptions of each of these measures are presented next. It should be noted, however, that these measures do not constitute a complete "mapping" of these domains since additional measures are planned for use at the second and third grade.

Descriptions of Measures

Measures of Academic Skills and Abilities

The seven measures of academic skills and cognitive-language abilities listed in Figure 2 are obtained from three tests, from the Parent Interview and from school records.

PIAT Word Recognition. The Word Recognition subtest of the Peabody Individual Achievement Test (PIAT) was selected for administration at the kindergarten level in spring and was continued in 1979. Six standard reading tests and several individually administered oral reading tests were considered before arriving at this decision.[1] The selection process was focused on "reading achievement" in general, so this discussion reviews the considerations involved in deciding upon tests to use from kindergarten through third grade, focusing on reading recognition at early grades and reading comprehension at second and third grades.

[1]Two widely used tests were not reviewed because they were considered inappropriate: the Iowa Test of Basic Skills starts at too high a level--grade 3; the Wide Range Achievement Test consists only of a list of words that must be correctly pronounced but cannot be reliably "sounded out."

Figure 2

PDC First Grade Child Measurement Battery
Spring 1979

Domain and Construct	Type of Measure[1]						
	Test	Inter-view	Teacher Rating	Tester Rating	Parent Rating	Records	Obser-vation
ACADEMIC SKILLS AND ABILITIES							
Reading Recognition	PIAT						
Mathematics	PIAT						
Verbal Fluency	MSCA						
Syntactic Development	BSM-English BSM-Spanish						
Progress in School					PI		
Retention in Grade						Records	
Special Education Placement						Records	
SOCIAL-EMOTIONAL DEVELOPMENT							
Social Problem-Solving	PIPS						
Attitude Toward Teacher		CI					
Self-Assurance			CRS-1				
Aggressiveness			CRS-2				
Dependence			CRS-3				
Sociability				POCL-2			
LEARNING ATTITUDES							
Attitude Toward Reading		CI					
Academic Motivation			CRS-4				
Task Orientation				POCL-1			
Attitude Toward School					PI		

[1]Key to measures:
PIAT: Peabody Individual Achievement Test
MSCA: McCarthy Scales of Children's Abilities
BSM: Bilingual Syntax Measure
PI: Parent Interview
PIPS: Preschool Interpersonal Problem Solving Test
CI: Child Interview
CRS: Child Rating Scale
POCL: Pupil Observation Checklist
FO: Focused Observation
CAR: Classroom Activities Record

(continued)

Figure 2

(continued)

Domain and Construct	Type of Measure[1]						
	Test	Inter-view	Teacher Rating	Tester Rating	Parent Rating	Records	Obser-vation
LEARNING ATTITUDES (cont.)							
Engagement in School-Related Work at Home					PI		
Attendance						Records	
HEALTH AND NUTRITION							
Height for Age	Direct						
Weight for Height	Measures						
CLASSROOM BEHAVIOR[2]							
Child-Initiated Inter-actions with Teacher							FO & CAR
Attention to Learning							FO
Respect for Parents							CAR
Disruption							CAR
Noise Level							FO
Respect for Teacher/Aide							FO
Cooperation with Teacher							FO

[1]Key to measures: PIAT: Peabody Individual Achievement Test
MSCA: McCarthy Scales of Children's Abilities
BSM: Bilingual Syntax Measure
PI: Parent Interview
PIPS: Preschool Interpersonal Problem Solving Test
CI: Child Interview
CRS: Child Rating Scale
POCL: Pupil Observation Checklist
FO: Focused Observation
CAR: Classroom Activities Record

[2]Classroom behavior measures are obtained from the Classroom Observation System and represent classroom-level indices of children's behavior, rather than the child-level measures obtained in the other domains.

The oral reading tests typically ask the child to read passages aloud, whereupon the tester asks relatively open-ended comprehension questions. These tests were not selected for use in the PDC evaluation for three major reasons: (1) the passages are out-dated and biased both in their middle-class, suburban, two-parent family orientation and in the traditional sex roles assigned to characters; (2) the comprehension questions often test world knowledge and logical reasoning skills rather than comprehension of the passages *per se*; (3) the tests are most sensitive for children with average, or above average, reading abilities, thus leading us to hypothesize serious "floor" effects with the kindergarten PDC sample.

The Comprehensive Tests of Basic Skills was rejected due to peculiar test items with curriculum bias and middle-class, suburban orientation in comprehension items which appeared to introduce genuine bias. The Stanford Achievement Test was rejected because the comprehension items require specific world knowledge not typically acquired through experience--it resembles a social studies test. The California Achievement Test was rejected due to a strong curriculum bias. Many items tested subskills related to a particular approach to teaching reading and used jargon that would be unfamiliar to teachers and students in many programs. The Sequential Tests of Educational Progress was rejected. Although it contained some of the best paragraph comprehension items of any of the tests reviewed, it also included vocabulary and sentence comprehension items which were frequently unclear and/or strongly culturally biased.

The PIAT has two reading subtests: Word Recognition and Reading Comprehension. As noted earlier, the Word Recognition subtest was selected for administration at the kindergarten and first grade levels. It begins with letter recognition and moves to a WRAT-like word list. Although the word list seems inappropriate, like the WRAT, the early items seemed satisfactory and thus suitable for kindergarten and first grade but not for second or third-grade. The Reading Comprehension subtest was rejected for use in the PDC evaluation. It involves single sentences which are read silently and matched to line drawings. It seems to have an unacceptably large memory component and some items are dated or culturally biased.

The PIAT Word Recognition subtest is individually administered and has 84 items ranging in difficulty from preschool through high school. Items 1-9 involve matching one or more letters with identical stimuli that must be discriminated by the child from other increasingly similar shapes in a multiple-choice format. Items 10-18 present individual letters to be named, and items 19-84 are individual words to be read aloud. The subtest can be administered in about 10 minutes to first graders.

This subtest was used in the Home Start follow-up study (Bache & Nauta, 1979) with results that confirm its usefulness as a program evaluation measure. The following table compares data from the standardization sample (Dunn & Markwardt, 1970), the Home Start follow-up study and the PDC evaluation:

Grade Level	Standardization		Home Start Follow-Up			PDC Evaluation[1]	
	Mean	*SD*	*Mean*	*SD*	*Alpha*	*Mean*	*SD*
Kindergarten	11.9	3.83	15.6	4.3	.91	15.1	4.43
First Grade	21.0	5.15	24.1	6.0	.92	22.8	5.73
Second Grade	28.7	6.79	32.1	8.8	.94	Not Applicable	

The publishers report one-month test-retest reliabilities for reading recognition of .81 for the kindergarten sample and .89 for the first grade sample (the internal consistency, alpha, reliabilities reported in the Home Start study are undoubtedly inflated due to the careful ordering of items on the test and the testing procedure of establishing basal and ceiling levels rather than administering a fixed number of items to each child).

A major advantage of the PIAT, that applies to both the Recognition and Mathematics subtests, is that items are arranged in ascending order of difficulty and individual starting and finishing points can be determined on a child-by-child basis. (The basal rule is 5 consecutive correct responses; the ceiling rule is 5 errors in any 7 consecutive responses.) This procedure assures a reasonable match between the difficulty of the test and the test-related capabilities of the children. The procedure also avoids potential psychometric problems associated with out-of-level testing.

There are other advantages of the PIAT mode of administration. The lay-out is perceptually appealing, with large line drawings on glossy paper and clear type, both features concentrating attention upon the question at hand. It is impossible for children to skip an item accidentally. The PIAT is an untimed, power test, yet the procedure of establishing a critical range of items for each child allows the test to be given in approximately 10 minutes. Instructions, questions, or directions may be repeated either when requested by the child or when no response is elicited. And, if children are working too quickly, the tester can attempt to modify the tempo. Guessing is encouraged. Rapport between tester and child should be greater than in a group test because it can be established on a personal level, and the tester can attempt to motivate the child

[1] From preliminary analysis of Spring 1979 analytic samples, PDC and comparison groups combined.

through use of praise and encouragement. All of these aspects of the PIAT should help eliminate performance differences favoring programs that spend a large amount of instructional time in fostering test-taking skills. Finally, the individually administered character of the PIAT makes possible more detailed monitoring of the behavior of each child in the test situation.

The Metropolitan Achievement Test (MAT) emphasizes comprehension in a way that results in little or no curriculum bias and high face-validity. Items are well-constructed, but the content is more relevant to middle-class children than to PDC children. Since the comprehension questions relate closely to the passages, the culture bias is not as severe as it otherwise might be, but it is there. One difficulty with the MAT, however, is that the grade-appropriate tests appear to be too difficult; too little variance is anticipated to allow detecting small but real program effects. Thus, it was recommended that the PRIMER level be used in the PDC evaluation at second grade (it was normed for K.5 to 1.4) and that Primary 1 (1 - 2) be used at third grade, although it was normed for 1.5 to 2.4 (2.5 to 3.4). Pilot testing of the MAT PRIMER was conducted with a subsample of first grade PDC children in 1979 to provide more guidance on the choice of level (and also to examine the relationship between reading recognition and reading comprehension scores).

PIAT Mathematics. In our search for a test of mathematics achievement, six standard commercial tests were given serious consideration. These included the California Achievement Test, Sequential Tests of Educational Progress, Comprehensive Tests of Basic Skills, Metropolitan Achievement Tests, Stanford Achievement Test and the PIAT. The following discussion explains why our review resulted in selection of the PIAT Mathematics subtest for PDC.

Four major criteria were considered: (1) whether the test in fact measures mathematics abilities (and not, for example, reading or memory); (2) absence of bias toward a particular curriculum; (3) clear instructions that would tend to minimize the advantage for test-wise or test-trained students; (4) a minimum of cultural bias in the content of the items.

The PIAT Mathematics subtest is individually administered with a total of 84 items. This, in combination with characteristics of the format and items, gives it a number of advantages. Most important is the fact that reading ability is minimally confounded with mathematics achievement/ability. The separation of reading from mathematics is accomplished by having the tester read each question aloud. The child does not see a representation of the question in print, although for some questions the child is shown either printed or pictorial representations of key facts or directions (presumably to reduce the demands of maintaining the content of the question in memory).

A distinctive feature of content covered by the PIAT Math test is the emphasis upon a mix of problems that demand application of frequently taught mathematical concepts (e.g., computation, quantitative aspects of currency and time, ordinality, cardinality, shape, and specific terms such as "double," "fifths," "youngest," etc.) in relatively universal situations. Although children are not permitted to use pencil and paper, the computations required are such that it seems reasonable to expect children to perform them mentally. The problems are constructed so as to minimize the influence of previous experience with particular formats and content.

The PIAT Math subtest was also used in the Home Start follow-up study. The following table compares data from the standardization sample, the Home Start study and the PDC evaluation:

Grade Level	Standardization		Home Start Follow-Up			PDC Evaluation[1]	
	Mean	*SD*	*Mean*	*SD*	*Alpha*	*Mean*	*SD*
Kindergarten	13:1	3:98	13:1	4:0	.83	13.2	3.94
First Grade	19.5	6.32	20.1	6.3	.90	18.7	6:05
Second Grade	29.2	8.11	30.1	9:6	:93	Not Applicable	
Third Grade	37:5	9:12	Not Applicable			Not Applicable	

The publishers report one-month test-retest reliabilities for math of .52 for the kindergarten sample, .83 for first grade, and :68 for third grade. In the Home Start follow-up study PIAT Math scores correlated .70 with PIAT Reading Recognition scores; in PDC the two subtests correlated .53 at kindergarten.

MSCA--Verbal Fluency. The Verbal Fluency subtest of the McCarthy Scales of Children's Abilities (McCarthy, 1972) has been included in the PDC battery from the beginning. It is one of the few general developmental measures that we judge suitable for continuing to administer as the children become older. In a very brief period (five minutes or less) an assessment of children's ability to recall information in conceptual categories is assessed. The child is asked to name as many members of the category as he or she can within 20 seconds. The categories are "animals," "things to eat," "people's names," and "toys."

[1]From preliminary analysis of the Spring 1979 analytic samples, PDC and comparison groups combined.

In the first phase of the evaluation the test was administered at all grade levels to a sample of 24-27 children at the Georgia site; data from that testing, along with data from the total PDC sample are as follows:

Grade Level	Cross-grade Data from Georgia		Total PDC English-Dominant Sample[1]		
	Mean	*Alpha*	*Mean*	*SD*	*Alpha*
Head Start	Not Applicable		9.3	5.73	.76
Kindergarten	14.1	.71	14.3	5.79	.68
First Grade	16.4	.75	16.2	6.05	
Second Grade	20.8	.79	Not Applicable		
Third Grade	21.8	.67	Not Applicable		

Bilingual Syntax Measure (BSM). The BSM (Burt, Dulay, & Hernandez Chavez, 1975) was also selected at the outset of PDC. It represented a well-developed measure of syntactic development that was suitable for both Spanish- and English-speaking children. The BSM assesses children's oral proficiency with English and/or Spanish grammatical structures, using cartoon-like pictures to elicit children's responses to the tester's questions. Simple questions, used with the colored cartoon pictures, provide a conversational setting for eliciting speech. An analysis of the child's response yields a score that can range from 0 to 18. Responses are written down verbatim by the tester so that further analyses can be carried out at a later time if desired.

All children have been administered the BSM-English, and children who show facility in Spanish are administered the BSM-Spanish as well. For the children who receive both versions, the order of administration is controlled so that during any single testing period half the children receive the Spanish version first and half receive the English version first.

During the first phase of the PDC evaluation alpha coefficients at different time points ranged from .82 to .88 for the English-dominant sample and from .76 to .96 for the Spanish-dominant children. There may be some concern with a ceiling effect by third grade--the mean score at kindergarten was 11.93 on the BSM-English for English-dominant children (SD = 3.09) and 13.13 on the BSM-Spanish for Spanish-dominant children (SD = 2.45)--but a new edition of the test may be available for use at

[1]From preliminary analysis of the Spring 1979 analytic samples, PDC and comparison groups combined.

second and third grades. Interim Report IX reported good stability across testing times (e.g., correlations of .75 for the BSM-English from fall to spring of the Head Start year and .61 from fall Head Start to spring kindergarten), so the BSM seems to be a likely candidate for demonstrating real longitudinal trends from Head Start through third grade for the PDC evaluation.

Progress in school. To provide another perspective on school achievement it was decided to ask parents to judge the progress of their child during the parent interview. Question 11m on the first-grade Parent interview asks the parent to rate whether their child "is learning a lot at school" on a scale from 1 ("definitely true") to 5 ("not at all true").

Special education and grade retention. Two items typically kept in school records and used in some educational evaluations are being documented for PDC, although guidelines for interpreting data from these records are not completely clear at this time. The extent to which a child is "held back" a grade may be a very gross indicator of the school's capacity to provide adequate educational programming for that child. This information will be collected so that the incidence of grade retention can be examined for possible program effects. A number of longitudinal studies of Head Start and other preschool programs (see Lazar, Hubbell, Murray, Roche, & Royce, 1977) have found that early intervention reduces the extent to which children are retained in grade.

Placing children in "special education" has also been an indicator that the children are not succeeding in school to the extent expected of them. Since this has been a key impact variable for some longitudinal studies (see Lazar, et al., 1977), it was decided to collect information on it for the PDC evaluation. Unfortunately, special education placements have become a very different matter than they were in the late 1960's and early 1970's. With the advent of PL 94-142 and heightened awareness on the part of school personnel of the need to provide appropriate educational experiences for handicapped children, placement data per se will provide ambiguous information: a higher rate for PDC could mean children with special needs are being appropriately provided for or that the program has somehow "created" more children with special needs. At any rate, this information will be available for the PDC and comparison samples for use either as impact data or as explanatory information.

In Spring 1979, testers at each site asked school personnel whether children in the sample had been provided with an individualized education program (IEP). If the answer was "yes," information was collected on the nature of the handicap and on the extent of mainstreaming, using categories developed for a national evaluation of mainstreaming in Head Start (Walters, Vogel, Brandis, & Thouvenelle, 1978).

Categories of handicapping conditions are the following:

Hard of hearing
Deaf
Speech impaired
Visually handicapped
Blind
Seriously emotionally disturbed
Specific learning disability
Orthopedically impaired
Other health impaired
Mentally retarded
Multi-handicapped
Other

The extent of mainstreaming was classified as follows:

Complete mainstreaming into a regular classroom, with no supportive assistance
Complete mainstreaming into a regular classroom, with supportive assistance provided by a specialist
Complete mainstreaming in a classroom where the regular classroom teacher and a special education teacher cooperatively work with all children in a team arrangement
Reverse mainstreaming in which non-handicapped children become part of a special education class
Partial mainstreaming where handicapped child is in a special education class but goes to mainstreamed classes for one or more regular classroom activities
[illegible] child is not mainstreamed and spends the [illegible] day in special education classes

Measures of Social-Emotional Development

The six constructs within the social-emotional domain are assessed by four instruments: one test (PIPS), a child interview, a teacher rating scale (CRS) and a tester rating scale (POCL).

Social problem-solving. The Preschool Interpersonal Problem-Solving Test (PIPS) (Shure & Spivack, 1974) was selected for use at the beginning of the PDC evaluation and has remained an important part of the battery at each testing period. The PIPS assesses the child's ability to generate alternative solutions to the problem of obtaining a toy from another child. Seven different toys are depicted and used as the basis for the problem. For example:

"Johnny has been playing with this truck for a long time and Jimmy wants a chance to play with it. But Johnny keeps on playing with it. Jimmy wants to play with the truck. What can he do to get a chance to play with the truck?"

Testers probe to elicit as many different types of solutions as possible to the seven problem situations. Thus far the PIPS has been scored for the number of different solutions offered by the child, so that the range of scores is only 0 to 7, since there are seven opportunities to offer new solutions.

The solutions offered by the child are coded according to clearly defined rules. First, each response is coded as a "solution" or a "non-solution." There are 25 solution categories and 3 non-solution categories.

There is some concern that the test sample may be approaching a ceiling given the current scoring system, since in the Home Start follow-up study kindergarten, first-grade and second-grade groups all averaged around 4.3 different solutions (with x standard deviations of 1.6). Thus far in PDC, PIPS scores are increasing appropriately with age. The mean scores at entering Head Start, spring Head Start and spring kindergarten children were 2.1, 2.7 and 3.8, respectively.

One possibility for increasing the sensitivity of the PIPS to the social problem-solving strategies of older children would be to develop scores for different types of solutions. Among the 15 solution categories, there appear to be ones that are conceptually related. Using Spring 1979 data, procedures will be explored for categorizing the solutions into groups with children receiving a score for the number (or proportion) of solutions offered in each group. One possible grouping is: (1) prosocial solutions, e.g., ask, say please, share; (2) trade/swap solutions, e.g., trade or bribe; (3) manipulative/deceptive solutions, e.g., trick, finagle, manipulate; (4) anger; (5) passive solutions, e.g., wait, plan for future; and (6) aggressive solutions, e.g., grab, physical attack.

The PIPS authors claim that their measure is relatively unconfounded with general ability beyond some minimum necessary to understand the task. PDC data confirm this. Although factor analysis of the kindergarten data showed that PIPS loaded on a factor that included measures of verbal fluency and verbal memory, indicating a strong verbal component to the test, its bivariate correlations with those measures were only moderate (e.g., .27 between PIPS and Verbal Fluency) and PIPS scores correlated very slightly with other measures (e.g., .08 with BSM-English, .12 with PIAT Reading Recognition and .10 with PIAT Mathematics). The PIPS does not lend itself to estimates of internal consistency reliability and PDC data indicate low stability from one testing time to another compared

with the stability of other measures in the battery: the fall-to-spring correlation for the Head Start year was .41 and the correlation on PIPS scores from fall Head Start to spring kindergarten was .20. A more sensitive scoring procedure might also improve the stability of the measure.

Attitude toward teacher. A child's attitude toward his or her teacher is seen by many as an important aspect of successful performance. Without debating whether this attitude may be one of the causes of higher achievement levels or one of the effects, many program operators would be pleased to note improvements in children's attitudes toward their teachers. The Purdue Social Attitude Scales, developed by Cicirelli (19) contains a scale labeled "attitude toward school" in which children are shown cartoon-story items. Since a stick-figure representing a teacher plays a prominent role in most of the items, we have chosen to label the scale, "attitude toward teacher."

For each item, the main character (given the child's name by the tester) is about to be engaged in some activity or is about to have something happen to him or her. For example:

"_____(child's name) is working at school. _____'s teacher comes over. She looks at _____'s work. Which one is the teacher's face?"

The child responds to each item by pointing to one of five faces, which range in expression from very happy to very sad. The child's score is the mean "rating" given to the ten items. In designing the PDC battery for Spring 1979, 8 of 10 items were incorporated as Part 1 of an instrument called the Child Interview; Part 2 of the Child Interview--items assessing the child's attitudes toward reading--is described in the section on learning attitudes.

The "attitude toward teacher" items were administered for the first time in PDC in Spring 1979 but had been administered as part of the total Purdue Social Attitude Scale in the Home Start follow-up study. In that evaluation, relatively high internal consistency was found for the entire set of 30 items (alpha = .87 at first grade); however, the internal consistency for the 10 item set, from which the 8 items of the PDC Child Interview were drawn, was considerably lower (alpha = .67). Counterbalancing that finding were other factors. First, Home Start data indicated that Purdue Social Attitude Scale score was not confounded with age--group means did not differ significantly from kindergarten through second grades. Second, obviously better measures of young children's attitudes toward school were not available when instrument selection decisions had to be made.

Child Rating Scale (CRS). Three social-emotional constructs have been assessed in PDC with the Child Rating Scale (a fourth scale of the CRS, academic motivation, was added in 1979; it is discussed under "learning attitudes." The CRS was originally developed in the first year of PDC to assess children's behavior in two areas of "social competence": interpersonal competence and task competence. Thirty-nine statements (e.g., cooperates and shares with others; shows self-confidence) were developed, relying heavily on the Bronson General Competence Rating Scale (Bronson, 1973). The statements are listed on a response sheet and teachers are asked to rate each child on each item using a five-point scale (1 = rarely; 5 = very frequently).

Prior to Spring 1979 data collection it was decided to refine the Child Rating Scale so as to reduce the number of items (thereby reducing the burden on teachers) and to add a scale to assess academic motivation. Using factor analysis and regression procedures (as described fully in Interim Report IX), the original 39 items were reduced to 17 which possessed a stable factor structure of three scales, each with an internal consistency reliability exceeding .75. The three scales have been labeled "self-assurance," "aggressiveness," and "dependence." An examination of the aggressiveness scale items, however, suggests that the scale may confound pro-social and anti-social aggression, so the possibility of separating these into two scales is being examined.

Sociability. A rating scale completed by testers was originally developed by High/Scope for use in the Home Start Evaluation (Deloria, Love, Gordon, Harvey, Hockman, Platt, Nauta & Springer, 1974). This scale, the Pupil Observation Checklist (POCL) had its roots in a longer, 25-item scale developed by High/Scope for use in its Follow Through evaluation work. The 10-item version used in the Home Start evaluation and the slightly modified 11-item version used at each testing time in PDC have consistently yielded two stable, highly reliable factors. One scale, "sociability," is classified here as a construct within the social-emotional domain. The second scale, "task orientation," will be discussed under the domain of learning attitudes.

The POCL consists of 11 7-point rating items which are completed by the tester immediately after the testing session with each child. The sociability scale consists of the following bipolar adjectives:

Shy-sociable
Quiet-talkative
Passive-active

Measures of Learning Attitudes

Learning attitudes have assumed increasing importance among educators. Several constructs related to this general domain have been identified for measurement in PDC, and a variety of methods are being used to obtain the measures. The six constructs listed in Figure 2 are obtained from five different procedures: a child interview, a teacher rating, a tester rating, a parent rating and school records.

Attitude toward reading. Part II of the Child Interview consists of an informal discussion between the tester and child in which the tester asks the child a number of questions in a conversational manner (e.g., "Do you like story books?" "Do you like to read stories to other people?" "When do you most like to read?" "What is your favorite story book?"). After the conversation is finished the tester completes six ratings, each on a scale from 1 to 5:

1.	Child rarely reads at school.	Child reads a lot in school.
2.	Child rarely reads outside of school.	Child reads a lot outside of school.
3.	Child reads only what he/she has to read; does not read for pleasure or for own information.	Child reads a great deal on own initiative for pleasure and for information.
4.	Reading is perceived only as a school activity.	Child perceives reading as an important activity outside of school also.
5.	There is little or no variety in the reading materials used in school.	Considerable variety in school reading, e.g., readers, stories, library materials.
6.	There is little or no variety in the reading materials used outside of school.	Considerable variety in reading materials used outside of ool, e.g., stories, newspapers ters, magazines.

This procedure was adapted from a process used by the High/Scope Follow Through evaluation in 1976-77 (Kittel, Tamor, Smith & Bond. 1977). After experimenting with a more structured questioning approach, the current method of guided unstructured discussion followed by tester ratings was adopted. Using the questions as a guide, the interviewer is free to probe and explore the topics with each child until there is enough information to complete the ratings.

Scales one and five describe, from the child's perspective, how much time is spent reading in school and what variety of reading occurs in school. Scale six indicates what variety of reading materials are present and what variety of reading activity occurs in the home. These three scales appear to be measures of context rather than impact. Consequently, they are not considered in evaluating child outcomes.

The three remaining scales--two, three, and four--were formed to be highly intercorrelated (coefficient + alpha = .94) and were summed to form a component variable reflecting children's "interest in reading."

Academic motivation. As mentioned earlier, three items were added to the Child Rating Scale in an attempt to assess this construct. Several recent studies have emphasized the importance of this variable. In the Perry Preschool longitudinal study, for example, two teacher rating scales were found to correlate fairly highly with measured school achievements. One was the "academic motivation" scale of the Pupil Behavior Inventory (developed by Vinter, Sarri, Vorwaller, Schafer, 1966) and the other was an "academic potential" scale from the Ypsilanti Rating Scale. This dimension has also been emphasized in the work of Bloom (1976) and Cross (1978). The academic motivation factor of the Pupil Behavior Inventory originally contained nine items; ratings on that factor correlated from .49 to .71 with concurrent scores on the California Achievement Test at grades 1, 2, and 3 (Weikart, Bond, McNeil, 1978). Three items from the Pupil Behavior Inventory were selected for PDC:

Is motivated toward academic performance
Completes assignments
Is alert and interested in school work

Task orientation. Eight items from the Pupil Observation Checklist, described above, comprise a "task orientation" scale. As with the sociability scale, this measure has remained a stable and reliable index across the PDC testing times. The eight bipolar items, rated by the tester on a 7-point scale, are as follows:

Resistive-cooperative
Indifferent-involved
Easily distracted-attentive
Nervous-relaxed
Needs urging-quick to respond
Prefers easy tasks-attempts difficult tasks
Gives up easily-keeps trying
Needs reassurance, praise, encouragement-realistically self-confident

Attitude toward school and engagement in school-related work at home. During the Parent Interview, developed for administering the first time in Spring 1979 (see Section II of this appendix), the parent is asked several questions that are designed to tap two possibly distinct aspects of learning attitudes. The first, a general attitude toward school, consists of four items. The first two are statements read to the parent to which he or she responds by indicating whether each is "definitely true" to "not at all true" along a 5-point scale. The questions are:

(child's name) loves school and enjoys being there

_____ feels that he/she is learning a lot in school

The third item is a question which asks (if there are books or magazines in the home) whether the child usually asks someone to read with him/her or whether someone usually offers to read to the child (on the assumption that a child with a more positive attitude toward a school activity such as reading will be more likely to take the initiative in seeking to read with someone else). The fourth item asks whether the child voluntarily does homework or if he/she had to be prodded into doing it (after the interviewer ascertains that the child actually has homework assignments).

The second scale, "engagement in school-related work at home" was adopted because of concern that school achievement measured only in the school setting might not capture program effects that more thoroughly permeate a child's life. In High/Scope's Follow Through evaluation, it was discovered that even when differences were not found between Follow Through and comparison children on the standard achievement measures, Follow Through children were more likely to engage in activities at home that were related to their school work (Bond, Smith & Kittel, 1976). In programs like PDC and Follow Through, with their emphasis on parent participation in children's learning, this variable can be an important indicator of the program's success. The specific questions employed in the PDC Parent Interview are the following:

Not counting reading he/she has to do for school, how often does (child) look at a book or magazine at home? Would you say:

_____every day?
_____several times a week?
_____about once a week?
_____2 or 3 times a month? or
_____once a month or less?

Not counting homework, does (child) ever do things like writing or drawing that he/she learned at school?

_____ yes, often
_____ yes, sometimes
_____ no

Attendance. School attendance has frequently been cited as a possible indicator of general attitudes of the child toward school. We recognize that the interpretation of such data, however, is ambiguous. But, combined with other information, such information might provide additional insights into the influence of PDC programs on children.

Measures on Health and Nutrition

One of the most difficult areas to assess in a large-scale evaluation is the physical status of children. Fine-grained medical and nutritional assessments are too expensive and time-consuming to be feasible, and yet the provision of health and nutritional services is one of the critical features of Head Start programs and figures prominantly in the PDC guidelines. For the PDC evaluation it was decided to (1) obtain direct measures by measuring height and weight, recognizing that this would provide an assessment of only extreme departures from normal growth; (2) obtain data at the program level on the provision of health and nutritional services, assuming that if the delivery of services can be documented one can assume improved status on the part of children receiving the services; and (3) explore the feasibility of assessing the impact of the program on children's knowledge of good health and nutritional practices. As of Spring 1979, only the first of these procedures has been fully implemented.

Height and Weight. Each spring, High/Scope testers measure the children's height and weight using a standard procedure. Weight is measured on balance beam scales, usually borrowed from the school. A standard 10-pound weight is placed on the scale by the tester before weighing the children in order to check the calibration. Children were asked to remove their shoes and any outdoor clothing, such as jackets, before being weighed. Height is measured by having the child stand straight against a wall. The variables used in the analysis are (1) height for age, which is a good index of chronic protein undernutrition and (2) weight for height, an index of more acute changes in nutritional status.

Measures of Classroom Behavior

From the outset, the PDC evaluation has been concerned with the problem of inferring social-emotional variables from data collected in test-like settings and so a child-oriented observation system was developed and used from 1976 to 1978. Unfortunately, the costs involved in obtaining data that would be reliable at the individual child level became too prohibitive, and the information yield too ambiguous, for the procedure to be continued. In designing the second phase of the evaluation, it became clear that the evaluation needed to provide ACYF with more information on program implementation at the classroom level. Therefore, a new observation system was developed to describe PDC and comparison classrooms (see Section IV of this Appendix for a full description). Although the focus of that system is on program implementation, the observations necessarily involve children on a number of dimensions. Thus, to the extent that it is feasible given the response frequencies of the variables of interest, observation data can be analyzed at the classroom level to provide additional information about the influence of PDC on children's behavior.

The observation system includes seven variables that may be useful for this aspect of the evaluation. These are listed here, but the reader is referred to the PDC Classroom Observation System Manual (Diamondstone, Smith, & Rosario, 1979) for a more detailed explanation:

Child initiated interactions with teacher
Children's attention to learning activities
Children's respect for parents in the classroom
Level of classroom disruption
Noise level
Respect for teacher and aide
Cooperation with teacher

References

Anderson, S., & Messick, S. Social competency in young children. Developmental Psychology, 1974, 10 282-293.

Bache, W., & Nauta, M. J. Home Start Follow-up Study: A study of long-term impact of Home Start on program participants. Final Report to ACYF. Cambridge, MA: Abt Associates, Inc., 1979.

Bloom, B. S. Human characteristics and school learning. New York: McGraw-Hill, 1976.

Bond, J. T., Smith, A. G., & Kittel, J. M. Evaluation of curriculum implementation and child outcomes. Report to Office of Education, Project Follow Through. High/Scope Educational Research Foundation, 1976.

Bronson, M. Manual for executive and social skill profile for preschool children. Cambridge, MA: Harvard University, Graduate School of Education, 1973.

Burt, M., Dulay, H., & Hernandez-Chavez, E. Bilingual Syntax Measure. New York: Harcourt, Brace, Jovanovich, 1975.

Cicerelli, V. The Purdue Social Attitudes Scales. West Lafayette, Indiana: Purdue University, n.d.

Deloria, D., Love, J. M., Gordon, S., Hanvey, R., Hockman, E., Platt, J., Nauta, M., & Springer, N. The National Home Start Evaluation Interim Report V: Summative Evaluation Results. Report to ACYF. Ypsilanti, MI: High/Scope Educational Research Foundation, 1974.

Diamondstone, J., Rosario, J., & Smith, A. G. Project Developmental Continuity Classroom Observation System. Ypsilanti, MI: High/Scope Educational Research Foundation, 1979.

Dunn, L. M., & Markwardt, F. C., Jr. Manual for the Peabody Individual Achievement Test. Circle Pines, MN: American Guidance Service, 1970.

Kittel, J. M., Tamor, L., Smith, A. G., & Bond, J. T. Evaluation of curriculum implementation and its effects on children's written and oral language, and an exploratory study of post-treatment effects in LeFlore County, Mississippi. Report to Office of Education, Project Follow Through. High/Scope Educational Research Foundation.

Lazar, I., Hubbell, V. R., Murray, H., Rosche, M., & Royce, J. The persistence of preschool effects: A long-term follow-up of fourteen infant and preschool experiments. Report to ACYF. Ithaca, NY: Cornell University, 1977.

McCarthy, D. McCarthy Scales of Children's Abilities: Manual. New York: Psychological Corporation, 1972.

Ogilvie, D., & Shapiro, B. Manual for assessing social abilities of one to six year old children. In B. L. White & J. C. Watts (Eds.), Experience and environment: Major influences on the development of the young child. Englewood Cliffs, NJ: Prentice-Hall, 1969.

Raizen, S., & Bobrow, S. B. Design for a national evaluation of social competence in Head Start children. Santa Monica, CA: Rand Corporation, 1974.

Shure, M. B., & Spivack, G. The PIPS Test Manual. Philadelphia: Hahneman Medical College, 1974.

Vinter, R. D., Sarri, R. C., Vorwaller, D. J., & Schaefer, W.E. Pupil Behavior Inventory: A manual of administration and scoring. Ann Arbor, MI: Campus Publishers, 1966.

Walters, P. B., Vogel, R. J., Brandis, M. R., & Thouvenelle, S. Evaluation of the process of mainstreaming handicapped children into project Head Start. Report to ACYF. Silver Spring, MD: Applied Management Sciences. 1978.

Weikart, D. P., Bond, J. T., & McNeil, J. T. The Ypsilanti Perry Preschool Project: Preschool years and longitudinal results through fourth grade. Monographs of the High/Scope Educational Research Foundation, No. 3, 1978.

White, B. L., & Watts, J. C. (Eds.). Experience and environment: Major influences on the development of the young child. Englewood Cliffs, NJ: Prentice-Hall, 1973.

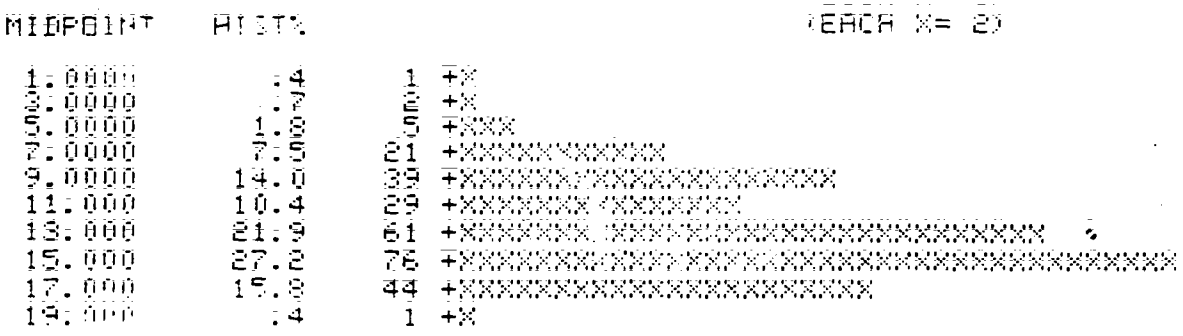
APPENDIX B

Comparison of Summary Score Distributions by Treatment Group, Child Outcome Measures for Spring 1979 (Grade 1)

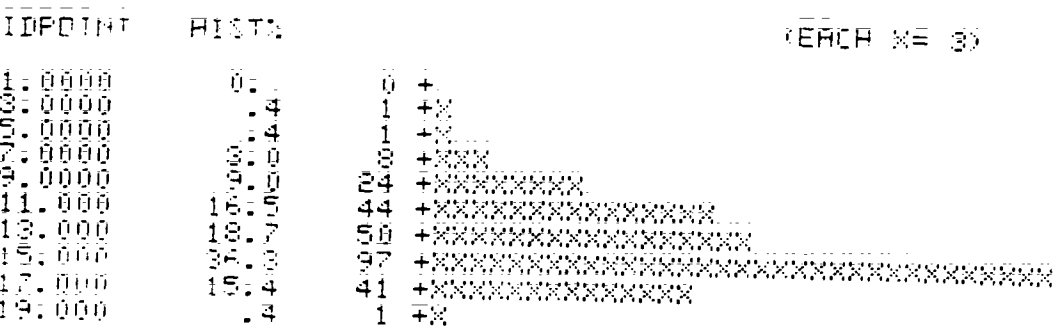
Figure 1

Comparison of Summary Score Distributions for PDC and Comparison Children on the Spring 1979 BSM-English

PDC *(n=279)*



Comparison *(n=267)*



Distributions significantly different: $\chi^2 = 19.97$, 9 d.f.; $p = .019$

Score levels at which significant differences occ[illegible]

Midpoint	Raw Score Range	Range in s.d. Units From Overall Mean	Direction of [illegible]ative Frequencies
7	6 - 7 (5.5 - 7.5)	-2.32 to -1.67	PDC > Comparison
9	8 - 9 (7.5 - 9.5)	-1.[illegible]7 to -1.00	PDC > Comparison
11	10 - 11 (9.5 - 11.5)	-1.00 to -0.34	PDC < Comparison

[1]Differences at $p \leq .10$ by Goodman's intervals test.

Figure 2

Comparison of Summary Score Distributions for PDC and Comparison Children on the Spring 1979 BSM-Spanish

PDC *(n=46)*

(EACH X= 1)

MIDPOINT	HIST%	
0.	2.2	1 +X
2.0000	0.	0 +
4.0000	0.	0 +
6.0000	4.3	2 +XX
8.0000	4.3	2 +XX
10.000	0.	0 +
12.000	4.3	2 +XX
14.000	32.6	15 +XXXXXXXXXXXXXXX
16.00	34.8	16 +XXXXXXXXXXXXXXXX
18.000	17.4	8 +XXXXXXXX

(INTERVAL WIDTH= 2.0000)

Comparison *(n=46)*

(EACH X= 1)

MIDPOINT	HIST%	
0.	2.2	1 +X
2.0000	0.	0 +
4.0000	0.	0 +
6.0000	0.	0 +
8.0000	0.	0 +
10.000	2.2	1 +X
12.000	15.2	7 +XXXXXXX
14.000	34.8	16 +XXXXXXXXXXXXXXXX
16.000	41.3	19 +XXXXXXXXXXXXXXXXXXX
18.000	4.3	2 +XX

(INTERVAL WIDTH= 2.0000)

Distributions not significantly different between groups.

Figure 3

Comparison of Summary Score Distributions for PDC and Comparison Children on the Spring 1979 MSCA Verbal Fluency

PDC *(n=279)*

MIDPOINT	HIST%		(EACH X= 3)
2.0000	1.4	4	+XX
7.0000	11.1	31	+XXXXXXXXXX
12.000	34.8	97	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
17.000	28.3	79	+XXXXXXXXXXXXXXXXXXXXXXXXX
22.000	15.4	43	+XXXXXXXXXXXXXX
27.000	4.7	13	+XXXX
32.000	3.2	9	+XXX
37.000	1.1	3	+X

(INTERVAL WIDTH= 5.0000)

Comparison *(n=267)*

MIDPOINT	HIST%		(EACH X= 3)
2.0000	.4	1	+X
7.0000	7.9	21	+XXXXXXX
12.000	31.1	83	+XXXXXXXXXXXXXXXXXXXXXXXXXXX
17.000	34.8	93	+XXXXXXXXXXXXXXXXXXXXXXXXXXXX
22.000	16.5	44	+XXXXXXXXXXXXXX
27.000	7.1	19	+XXXXXX
32.000	1.1	3	+X
37.000	.7	2	+X
42.000	.4	1	+X

(INTERVAL WIDTH= 5.0000)

Distributions not significantly different between groups.

Figure 4

Comparison of Summary Score Distributions for PDC and Comparison Children on the Spring 1979 PIAT-Mathematics

PDC *(n=274)*

(EACH X= 3)

MIDPOINT	HIST%	
6.0000	0.	0 +
11.000	21.9	60 +XXXXXXXXXXXXXXXXXXXX
16.000	37.6	103 +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
21.000	22.6	62 +XXXXXXXXXXXXXXXXXXX
26.000	12.0	33 +XXXXXXXXXX
31.000	3.3	9 +XXX
36.000	2.2	6 +XX
41.000	.4	1 +X

(INTERVAL WIDTH= 5.0000)

Comparison *(n=261)*

(EACH X= 3)

MIDPOINT	HIST%	
6.0000	.8	2 +X
11.000	15.7	41 +XXXXXXXXXXXXXX
16.000	35.2	92 +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
21.000	31.4	82 +XXXXXXXXXXXXXXXXXXXXXXXXXXX
26.000	9.2	24 +XXXXXXXX
31.000	3.8	10 +XXXX
36.000	2.7	7 +XXX
41.000	1.1	3 +X

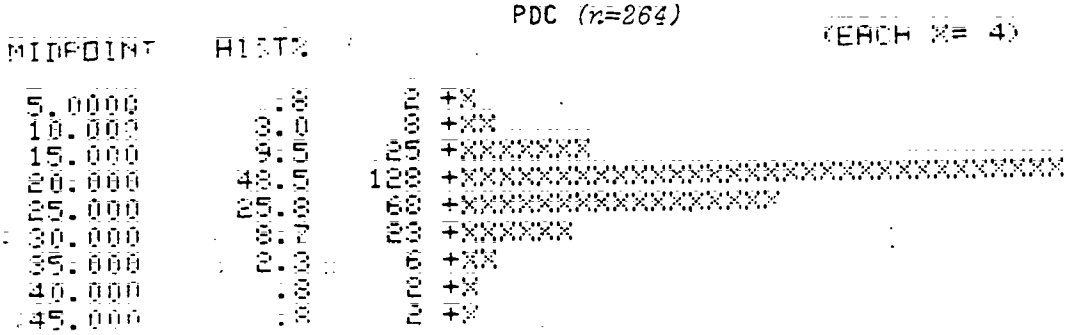
(INTERVAL WIDTH= 5.0000)

Distributions not significantly different between groups.

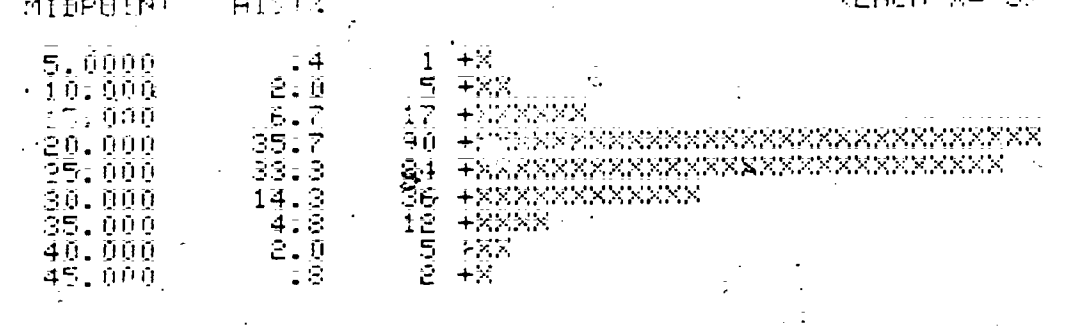
Figure 5

Comparison of Summary Score Distributions for PDC and Comparison Children on the Spring 1979 PIAT-Reading

PDC *(n=264)*



Comparison *(n=252)*



Distributions significantly different: $\chi^2 = 16.74$, 8 d.f.; p = .03

Score levels at which significant differences occur[1]:

Midpoint	Raw Score Range	Range in s.d Unit from Overall Mean	Direction of Relative Frequencies
20.00	18 - 22 (17.5 - 22.5)	-0.93 to -.06	PDC > Comparison

[1]Differences at p ≤ .10 by Goodman's intervals test.

Figure 6

Comparison of Summary Score Distributions for PDC and Comparison Children on the Spring 1979 PIPS

PDC *(n=279)*

```
MIDPOINT    HIST%                              (EACH X= 2)

 0.            .7      2 +X
 1.0000       7.9     22 +XXXXXXXXXXX
 2.0000       9.0     25 +XXXXXXXXXXXXX
 3.0000      16.1     45 +XXXXXXXXXXXXXXXXXXXXXXX
 4.0000      20.1     56 +XXXXXXXXXXXXXXXXXXXXXXXXXXX
 5.0000      20.4     57 +XXXXXXXXXXXXXXXXXXXXXXXXXXXX
 6.0000      17.2     48 +XXXXXXXXXXXXXXXXXXXXXXX
 7.0000       6.8     19 +XXXXXXXXX
 8.0000       1.8      5 +XXX

                    (INTERVAL WIDTH= 1.0000)
```

Comparison *(n=266)*

```
MIDPOINT    HIST%                              (EACH X= 2)

 0.            .4      1 +X
 1.0000       4.5     12 +XXXXXX
 2.0000       9.8     26 +XXXXXXXXXXXXX
 3.0000      21.1     56 +XXXXXXXXXXXXXXXXXXXXXXXXXXX
 4.0000      21.4     57 +XXXXXXXXXXXXXXXXXXXXXXXXXXXX
 5.0000      22.9     61 +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 6.0000      15.8     42 +XXXXXXXXXXXXXXXXXXXX
 7.0000       3.8     10 +XXXXX
 8.0000        .4      1 +X

                    (INTERVAL WIDTH= 1.0000)
```

Distributions not significantly different between groups.

Figure 7

Compar[illegible] of Summary Score Distributions for PDC and Comparison [illegible]en on the Spring 1979 POCL 1 "Task Orientation"

PDC *(n=259)*

MIDPOINT	[illegible]		(EACH X= 2)
11.000	[illegible]	0	+
16.000	[illegible]	5	+XXX
21.000	[illegible]	5	+XXX
26.000	[illegible]	19	+XXXXXXXXXX
31.000	16.[illegible]	42	+XXXXXXXXXXXXXXXXXXXX
36.000	12.[illegible]	33	+XXXXXXXXXXXXXXXXX
41.000	12.[illegible]	33	+XXXXXXXXXXXXXXXXX
46.000	18.[illegible]	47	+XXXXXXXXXXXXXXXXXXXXXXXX
51.000	12.4	32	+XXXXXXXXXXXXXXXX
56.000	16.[illegible]	42	+XXXXXXXXXXXXXXXXXXXX

(INTERVAL WIDTH= 5.0000)

Comparison *(n=254)*

MIDPOINT	HIST%		(EACH X= 2)
11.000	.8	[illegible]	+[illegible]
16.000	1.2	[illegible]	+[illegible]
21.000	1.2	[illegible]	+[illegible]
26.000	6.7	17	+[illegible]
31.000	13.0	33	+[illegible]
36.000	14.6	37	+[illegible]
41.000	15.4	[illegible]	+[illegible]
46.000	12.2	31	[illegible]
51.000	16.5	42	[illegible]
56.000	17.7	45	[illegible]

(INTERVAL WIDTH= 5.0000)

Distributions not significantly different between groups.

Figure 8

Comparison of Summary Score Distributions for PDC and Comparison Children on the Spring 1979 POCL 2 "Sociability"

PDC *(n=259)*

MIDPOINT	HIST%	(EACH X= 2)
3.0000	1.2	3 +XX
6.0000	6.6	17 +XXXXXXXXX
9.0000	13.1	34 +XXXXXXXXXXXXXXXXX
12.000	21.2	55 +XXXXXXXXXXXXXXXXXXXXXXXXXXX
15.000	22.4	58 +XXXXXXXXXXXXXXXXXXXXXXXXXXXX
18.000	14.7	38 +XXXXXXXXXXXXXXXXXX
21.000	20.8	54 +XXXXXXXXXXXXXXXXXXXXXXXXXX

(INTERVAL WIDTH= 3.0000)

Comparison *(n=254)*

MIDPOINT	HIST%	(EACH X= 2)
3.0000	.4	1 +X
6.0000	5.9	15 +XXXXXXXX
9.0000	8.3	21 +XXXXXXXXXX
12.000	22.0	56 +XXXXXXXXXXXXXXXXXXXXXXXXXXX
15.000	16.1	41 +XXXXXXXXXXXXXXXXXXXX
18.000	26.4	67 +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
21.000	20.9	53 +XXXXXXXXXXXXXXXXXXXXXXXXXX

(INTERVAL WIDTH= 3.0000)

Distributions significantly different: $\chi^2 = 15.10$, 6 d.f.; $p = .020$

Score levels at which significant differences occur[1]:

Midpoint	Raw Score Range	Range in s.d. Units From Overall Mean	Direction of Relative Frequencies
18	17 - 19 (16.5 - 19.5)	+0.33 to +0.98	Comparison > PDC

[1]Differences at $p \leq .10$ by Goodman's intervals test.

Figure 9

Comparison of Summary Score Distributions for PDC and Comparison Children on the Spring 1979 CRS-1 "Self-Assurance"

PDC *(n=264)* (EACH X= 2)

MIDPOINT	HIST%		
11.000	.4	1	+X
16.000	.8	2	+X
21.000	4.9	13	+XXXXXXX
26.000	11.4	30	+XXXXXXXXXXXXXXX
31.000	24.2	64	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
36.000	27.7	73	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
41.000	15.2	40	+XXXXXXXXXXXXXXXXXXX
46.000	9.1	24	+XXXXXXXXXXX
51.000	3.8	10	+XXXXX
56.000	2.7	7	+XXXX

(INTERVAL WIDTH= 5.0000)

Comparison *(n=246)* (EACH X= 3)

MIDPOINT	HIST%		
11.000	0.	0	+
16.000	1.2	3	+X
21.000	2.4	6	+XX
26.000	12.2	30	+XXXXXXXXXX
31.000	34.6	85	+XXXXXXXXXXXXXXXXXXXXXXXXXXXX
36.000	25.6	63	+XXXXXXXXXXXXXXXXXXXXX
41.000	9.8	24	+XXXXXXXX
46.000	8.5	21	+XXXXXXX
51.000	5.3	13	+XXXXX
56.000	.4	1	+X

(INTERVAL WIDTH= 5.0000)

Difference between distributions at near-significant levels: $\chi^2 = 15.95$, 9 d.f.; $p = .068$ [Maximum likelihood test: 17.00, $p = .049$]

Score levels at which significant differences occur[1]:

Midpoint	Raw Score Range	Range in s.d. Units From Overall Mean	Direction of Relative Frequencies
31.00	29 - 33 (28.5 - 33.5)	-0.89 to -0.22	Comparison > PDC
41.00	39 - 43 (38.5 - 43.5)	+0.44 to +1.11	PDC > Comparison
56.00	54 - 55 (53.5 - 55.0)	+2.44 to +2.64	PDC > Comparison

[1]Differences at $p \leq .10$ by Goodman's intervals test.

Figure 10

Comparison of Summary Score Distributions for PDC and Comparison Children on the Spring 1979 CRS-2 "Aggressiveness"

PDC *(n=267)*

```
MIDPOINT    HIST%                                  (EACH X= 2)

 4.0000      6.4     17 +XXXXXXXXX
 6.0000     12.7     34 +XXXXXXXXXXXXXXXXX
 8.0000     18.7     50 +XXXXXXXXXXXXXXXXXXXXXXXXX
10.000      20.2     54 +XXXXXXXXXXXXXXXXXXXXXXXXXXX
12.000      16.5     44 +XXXXXXXXXXXXXXXXXXXXX
14.000       9.7     26 +XXXXXXXXXXXXX
16.000      10.9     29 +XXXXXXXXXXXXXX
18.000       3.0      8 +XXXX
20.000       1.9      5 +XXX

                  (INTERVAL WIDTH= 2.0000)
```

Comparison *(n=254)*

```
MIDPOINT    HIST%                                  (EACH X= 2)

 4.0000      7.1     18 +XXXXXXXXX
 6.0000     11.4     29 +XXXXXXXXXXXXXX
 8.0000     16.1     41 +XXXXXXXXXXXXXXXXXXXX
10.000      18.9     48 +XXXXXXXXXXXXXXXXXXXXXXXX
12.000      23.2     59 +XXXXXXXXXXXXXXXXXXXXXXXXXXXX
14.000      15.0     38 +XXXXXXXXXXXXXXXXXX
16.000       3.9     10 +XXXXX
18.000       3.5      9 +XXXXX
20.000        .8      2 +X

                  (INTERVAL WIDTH= 2.0000)
```

Distributions significantly different: $\chi^2 = 16.39$, 8 d.f.; $p = .038$

Score levels at which significant differences occur[1]:

Midpoint	Raw Score Range	Range in s.d. Units From Overall Mean	Direction of Relative Frequencies
16.00	15 - 16 (14.5 - 16.5)	+1.23 to +1.78	PDC > Comparison

[1]Difference at $p \leq .10$ by Goodman's intervals test.

Figure 11

Comparison of Summary Score Distributions for PDC and Comparison Children on the Spring 1979 CRS-3 "Dependence"

PDC *(n=266)*

MIDPOINT	HIT%	(EACH X= 2)
2.0000	8.6	23 +XXXXXXXXXXX
3.0000	6.0	16 +XXXXXXXX
4.0000	14.3	38 +XXXXXXXXXXXXXXXXXX
5.0000	12.8	34 +XXXXXXXXXXXXXXXX
6.0000	24.4	65 +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
7.0000	16.2	43 +XXXXXXXXXXXXXXXXXXXX
8.0000	12.4	33 +XXXXXXXXXXXXXXXX
9.0000	3.8	10 +XXXXX
10.000	1.5	4 +XX

(INTERVAL WIDTH= 1.0000)

Comparison *(n=256)*

MIDPOINT	HIT%	(EACH X= 2)
2.0000	12.1	31 +XXXXXXXXXXXXXXX
3.0000	7.8	20 +XXXXXXXXXX
4.0000	11.3	29 +XXXXXXXXXXXXXX
5.0000	16.0	41 +XXXXXXXXXXXXXXXXXXX
6.0000	31.3	80 +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
7.0000	10.2	26 +XXXXXXXXXXXX
8.0000	6.6	17 +XXXXXXXX
9.0000	2.7	7 +XXXX
10.000	2.0	5 +XXX

(INTERVAL WIDTH= 1.0000)

Distributions not significantly different between groups.

Figure 12

Comparison of Summary Score Distributions for PDC and Comparison Children on the Spring 1979 CRS-4 "Academic Motivation"

PDC *(n=264)*

MIDPOINT	HIST%		(EACH X= 2)
3.0000	2.7	7	+XXXX
5.0000	5.7	15	+XXXXXXXX
7.0000	12.1	32	+XXXXXXXXXXXXXXXX
9.0000	28.8	76	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
11.000	15.2	40	+XXXXXXXXXXXXXXXXXXXX
13.000	19.3	51	+XXXXXXXXXXXXXXXXXXXXXXXXX
15.000	16.3	43	+XXXXXXXXXXXXXXXXXXXXXX

(INTERVAL WIDTH= 2.0000)

Comparison *(n=253)*

MIDPOINT	HIST%		(EACH X= 3)
3.0000	4.0	10	+XXXX
5.0000	6.7	17	+XXXXXX
7.0000	13.0	33	+XXXXXXXXXX
9.0000	34.4	87	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXX
11.000	15.4	39	+XXXXXXXXXXXX
13.000	13.0	33	+XXXXXXXXXX
15.000	13.4	34	+XXXXXXXXXX

(INTERVAL WIDTH= 2.0000)

Distributions not significantly different between groups.

Figure 13

Comparison of Summary Score Distributions for PDC and Comparison Children on the Spring 1979 CI-1 "Attitude Toward Teacher"

PDC *(n=279)*

MIDPOINT	HIST%		(EACH X= 4)
1.0000	10.8	30	+XXXXXXXX
2.0000	48.0	134	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
3.0000	35.5	99	+XXXXXXXXXXXXXXXXXXXXXXXX
4.0000	5.4	15	+XXXX
5.0000	.4	1	+X

(INTERVAL WIDTH= 1.0000)

Comparison *(n=265)*

MIDPOINT	HIST%		(EACH X= 3)
1.0000	12.1	32	+XXXXXXXXXX
2.0000	42.6	113	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
3.0000	41.9	111	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
4.0000	3.4	9	+XXX
5.0000	0.	0	+

(INTERVAL WIDTH= 1.0000)

Distributions not significantly different between groups.

Figure 14

Comparison of Summary Score Distributions for PDC and Comparison Children on the Spring 1979 CI-2 "Interest in Reading"

PDC *(n=272)*

MIDPOINT	HIST%	(EACH X= 2)
3.0000	7.4	20 +XXXXXXXXXX
5.0000	5.1	14 +XXXXXXX
7.0000	13.2	36 +XXXXXXXXXXXXXXXXXX
9.0000	27.6	75 +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
11.000	17.3	47 +XXXXXXXXXXXXXXXXXXXXXXX
13.000	21.7	59 +XXXXXXXXXXXXXXXXXXXXXXXXXXXX
15.000	7.7	21 +XXXXXXXXXX

(INTERVAL WIDTH= 2.0000)

Comparison *(n=257)*

MIDPOINT	HIST%	(EACH X= 2)
3.0000	6.2	16 +XXXXXXXX
5.0000	5.1	13 +XXXXXXX
7.0000	16.7	43 +XXXXXXXXXXXXXXXXXXXXX
9.0000	22.2	57 +XXXXXXXXXXXXXXXXXXXXXXXXXX
11.000	15.6	40 +XXXXXXXXXXXXXXXXXX
13.000	23.3	60 +XXXXXXXXXXXXXXXXXXXXXXXXXXXX
15.000	10.9	28 +XXXXXXXXXXXXX

(INTERVAL WIDTH= 2.0000)

Distributions not significantly different between groups.