DOCUMENT RESUME

ED 221 308 PS 013 094

AUTHOR Granville, Arthur C.; And Others
TITLE Assessment of Program Impact through the Kindergarten Year. Evaluation of Project Developmental Continuity. Interim Report IX.
INSTITUTION High/Scope Educational Research Foundation, Ypsilanti, Mich.
SPONS AGENCY Office of Child Development (DHEW), Washington, D.C. Early Childhood Research and Evaluation Branch.
PUB DATE Oct 79
CONTRACT HEW-105-78-1307
NOTE 108p.; For related documents, see ED 144 715, ED 160 235-248, ED 205 275-280, and PS 013 093.

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS Attrition (Research Studies); Classroom Observation Techniques; Comparative Analysis; Early Childhood Education; Interpersonal Competence; *Kindergarten Children; *Measures (Individuals); *Participant Characteristics; *Program Effectiveness; Spanish Speaking; Tables (Data)
IDENTIFIERS Impact Studies; *Project Developmental Continuity; Project Head Start

ABSTRACT

Presented in this report are results from a study conducted to evaluate the impact of Project Developmental Continuity on children's development and academic achievement. Following the first chapter's introductory overview of the study's objectives and organization, chapter 2 documents the data collection methods and summarizes the analysis questions addressed in the remainder of the report. Chapter 3 presents the results of analyses designed to highlight the nature of the sample and the characteristics of the various instruments, including a presentation of descriptive data taken on the sample and an analysis of attrition patterns and the effects of attrition on the characteristics of the sample. Information about the instruments is then presented in terms of the distributions of scores, internal consistency and stability of individual measures, the factor structure of the battery, and the relationships of the measures to criteria of "social competence." Chapter 4 examines treatment effects, presenting findings of program impact separately for English-dominant and Spanish-dominant samples. Finally, all results are summarized in chapter 5. A description of the instruments used in the study is appended. (MP)

EVALUATION OF PROJECT DEVELOPMENTAL CONTINUITY

INTERIM REPORT IX

ASSESSMENT OF PROGRAM IMPACT THROUGH THE KINDERGARTEN YEAR

October 1979

Prepared by:

Arthur C. Granville
Mary Morris
John M. Love
Nancy Hepler

With the Assistance of:

Jana von Fange
Cathy Peterson
Barb Bruemmer
Ann Hale
Jeffrey E. Moore
Linda Eckel

Submitted to the Early Childhood Research and Evaluation Branch, Administration for Children, Youth and Families, Office of Human Development Services, Department of Health, Education and Welfare, 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. This is an interim report to ACYF intended for use only by agency staff.

Table of Contents

Table of Contents
(continued)

Table of Contents
(continued)

List of Tables

List of Tables
(continued)

List of Figures

I

INTRODUCTION

Overview of the Evaluation

Project Developmental Continuity (PDC) began in 1974 as a concerted effort to demonstrate the value of coordinating Head Start and elementary school programs. Fifteen Head Start grantees were selected to establish administrative linkages between Head Start centers and elementary schools within their communities, and to design programs that would provide continuity of program and services. Twelve sites are still participating in the program and eleven are included in the evaluation.

The evaluation of this national demonstration effort was planned in two phases. In the first phase (1974 to 1978) a careful study was made to determine whether adequate conditions existed for carrying out a longitudinal study of the program's impact. This study would focus on a single cohort of children who began Head Start in fall 1976 and would complete third grade in the spring of 1981. Findings from the feasibility study--summarized in a final report (Love, Granville, & Smith, 1978) and detailed in Interim Report VII (Granville, Love, & Morris, 1977)--indicated that the child measures: (a) were sufficiently reliable and valid, (b) yielded scores that were sensitive to change beyond that which is simply a function of maturation, (c) related to teachers' and testers' judgments of the children's "social competence," (d) would be suitable for use with children in grades K-3, and (e) were generally suitable for administration by paraprofessional testers. The feasibility study also indicated that the evaluation sample was suitable in two main respects: (a) the PDC and comparison group samples appeared highly similar in terms of children's background characteristics and entering test performance; and (b) there was evidence from school records that there would be sufficient retention of the sample children over a five-year period so that longitudinal impact analyses could be performed.

A Preview of Plans for the Future Evaluation

This report represents a transition point in the PDC evaluation. The conclusions from the feasibility study were accepted by the Administration for Children, Youth and Families (ACYF) and the longitudinal phase of the evaluation was launched

in September 1978. Thus, this report on PDC's impact through the end of the kindergarten year (spring 1978) represents the status of the sample as the study enters its longitudinal phase. This is also a transition point in terms of the richness of information available about the various participants in PDC. At the Head Start and kindergarten levels our data collection for the impact study was limited to data on child outcomes. Beginning in 1979, in the spring of the children's first grade year, a more comprehensive data collection began, with parent interviews, teacher interviews, classroom observations, and administrator interviews added to the child measures.

The findings reported here raise more questions than they answer. We can describe the lack of clear evidence of the program's impact on those characteristics of children's development and achievement that are measured by the various tests, observations and ratings. We cannot, however, offer much in the way of satisfactory explanation for these findings. We do not know, for example, the extent to which PDC teachers really implemented an individualized instructional program, or the extent to which they differed from the comparison group teachers on this and other factors. We do not know the extent to which PDC parents participated in classroom activities with their children, or whether PDC parents did more of this than parents of comparison group children. The next report in this series, focusing on spring 1979 data, should provide answers to questions that seek explanations for the impact findings. It is our hope that the present report is useful in raising the appropriate questions for exploration as our analyses of the more comprehensive data begin.

This report follows the pattern established in preceding impact reports. Chapter II documents the methods of data collection and summarizes the analysis questions addressed in the remainder of the report. Chapter III presents the results of analyses designed to help us understand the nature of the sample and the characteristics of the various instruments. It includes a presentation of descriptive data on the sample and an analysis of attrition patterns and the effects of attrition on the characteristics of the sample. Chapter III then presents information on the instruments, in terms of the distributions of scores, internal consistency and stability of individual measures, the factor structure of the battery, and the relationships of the measures to criteria of "social competence." Chapter IV examines treatment effects, presenting findings of program impact separately for the English-dominant and Spanish-dominant samples. The entire report is summarized in Chapter V.

II

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 four 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 a detailed training manual was produced.
- A training model was designed that specified tester performance standards and provided for a four-day tester training session with large-group, small-group and individualized instruction, daily reviews of each tester's performance, and discussions of potential problems.
- Onsite monitoring of testers by trainers was conducted prior to the start of the actual testing.
- 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[1]

The roles of the personnel who conducted field data collection were explicitly defined in the Field Procedures Manual in order to systematize responsibilities. For example, site coordinator responsibilities included contacting the PDC coordinator regarding the start of testing; setting up and chairing a meeting with the kindergarten teachers involved in the evaluation, or contacting them individually; keeping in contact with High/Scope's supervisor of field operations about the status of data collection and any problems that the site was having; checking all completed data on a weekly basis; keeping up-to-date records on the status of the data collection; carrying out any needed training; testing (and, in some cases, observing) children; and monitoring testers.

Training Model

Training sessions for both trainers and testers were held in early April 1978. Since all but one of the seven trainers had been involved in previous PDC training sessions, a brief, one-and-one-half-day session was held to review and practice the child measures. More time was then spent with the new trainer on a one-to-one basis. In addition, the three observer trainers met for one day to review the observation system and code videotapes. The ten observers spent two days focusing on the PDC Classroom Observation System before being joined by the 17 testers for four days of training in the child measures.

Tester training. During the onsite tester training sessions, each test was reviewed and practiced. Practice sessions involved the use of test "scripts." The scripts 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 since 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.

---

[1]The spring data collection began the week of April 17 at all sites except Florida. The length of the data collection period was fairly constant across sites, with most testers finishing within seven or eight weeks. In Florida, testing and classroom observations were conducted over a three week period by High/Scope staff.

Since it is critical that testers administer the tests in a standard way, each tester was systematically "checked-out" on all of the child measures before the end of the training session. During this procedure a 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 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 the tester to say if the 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 could not be 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.

Observation training. As mentioned, the two days preceding the tester training session were devoted to observer training. In previous training sessions one tester from each site had been identified and trained as the PDC classroom observer. Thus, nine of the ten interviewers were familiar with the observation system. During this time the trainers reviewed the entire system, paying particular attention to those categories or items that the observers had had problems coding during the previous data collection period. The observers and trainers spent many hours coding videotapes and discussing the correct codes. The final training activity involved the coding of the observation reliability tape.

Monitoring

Onsite monitoring. The onsite monitoring occurred the week following the training session. At this time each of the newly hired testers was required to administer the PDC measures to a child while the trainer observed. After watching a tester, the trainer provided any additional feedback to the tester that was necessary for improving her interactions with the children. This procedure served two purposes: it gave the trainer an indication of how well the new testers could

establish rapport and interact with children, and it helped alleviate some of the anxieties the inexperienced testers felt about administering the measures to children.

Weekly monitoring. During the course of a testing week, testers alternately monitored each other; the one acting as monitor 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 the site coordinator at the end of each week. The site coordinator then checked these tests, plus any she had completed, for recording/scoring errors. (Site coordinators and interviewers had reviewed a checklist specifying what to look for when reviewing each completed booklet, e.g., "Is the identification complete?" "Did the interviewer fail to give a second trial when it should have been given?" "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 coordinator also kept track of all completed data (in addition to the individual records each tester kept of her classes) and mailed 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 were also checked by the supervisor of field operations at the High/Scope Foundation. Errors in recording or coding were identified and explained to the site coordinator, who then discussed them with the other testers.

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

Data Collection Sequence

Once the evaluation sample children were located in the district schools the testers 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 tester would be testing both PDC and comparison children, thus eliminating the possibility of tester bias for either group.

Determining Child's Language Capabilities

The procedure followed by testers in determining the language capabilities of children in the testing sample was (1) to ask the child's classroom teacher for his or her judgment, (2) to observe the child's verbal behavior in natural classroom conditions, and (3) on the basis of these indications, to administer the English or Spanish version, or both, of the Bilingual Syntax Measure (BSM). In most cases, this screening process produced consistent conclusions, and subsequent testing was accordingly conducted in English or Spanish or both. (In some cases this screening process led to the conclusion that a child was proficient in some third language, but not English or Spanish; these children were excluded from the testing sample.) When the screening process proved inconclusive, the tester carefully weighed all available information to reach a conclusion about the child's language capabilities.

Data Analysis Procedures

Chapters III and IV of this report present the results of four stages of analysis, focusing on:

- descriptive characteristics of the PDC and comparison samples
- attrition patterns within the samples
- characteristics of the instruments in the spring 1978 PDC battery

- effects of the PDC program on participating children.

The procedures for these analyses are described briefly below.

Descriptive Characteristics of the Samples

In order to characterize the composition of the PDC and comparison samples, descriptive statistics were computed for children in these samples at each site and at all sites combined. These statistics are based on the full sample of children tested, but subsequent analyses were carried out on an analytic subsample that excluded children who have handicaps likely to impair test performance or whose dominant language is not that of their local testing sample.

Attrition Patterns

Representativeness of remaining samples. Children who have departed the PDC and comparison samples since fall of their Head Start year were compared with the children remaining in the respective samples on a number of background variables and on fall 1976 test scores. The purpose of this phase of analysis was to determine whether any selection effect may be operating that diminishes the representativeness of the samples remaining. The hypothesis of bias due to attrition was evaluated by means of *t* tests and chi-square analyses.

Comparability of remaining samples. The remaining PDC and comparison samples were compared in the same way on the same dimensions to determine whether they may still be considered equivalent in terms of their characteristics at the time of program entry (fall 1976).

Characteristics of the Instruments

Because all the instruments in the battery have, in earlier PDC analyses, passed through at least two screenings based on psychometric criteria, no further screening was considered necessary before the data were entered into the analyses reported here. The psychometric analyses presented in this report are included mainly for purposes of instrument definition and documentation. The psychometric properties reported, on the basis of past and present analyses, include:

- means and standard deviations
- reliability (internal consistency)

- validity
- stability
- sensitivity to change
- intercorrelations
- factor structure
- relevance to "social competence."

Examination of Treatment Effects

In examining PDC's effects on children, three analytic questions were posed:

1. *Has a group difference [between PDC and comparison children] arisen since fall of the Head Start year?*
2. *Has a group difference arisen since spring of the Head Start year?*
3. *Is there a difference in the growth curves of the two groups?*

These questions were addressed by means of analysis of covariance and repeated measures analysis of covariance.

In separate analyses, PDC's effects at the classroom level (as distinguished from the child level) were examined on dimensions measured by the original version of the PDC Classroom Observation System. This version, now superseded by a new system, was used for the last time in spring 1978.

III

CHARACTERISTICS OF THE SAMPLE AND OF THE INSTRUMENTS

Characteristics of the Sample

General Description

A total of 782 children were tested at eleven PDC sites in spring 1978. Table 1 shows the number of children in the PDC and comparison groups at each site and describes the composition of each group.

Not all of the children tested entered into the analyses presented in this report; the actual number of children included in analyses is shown in the far right-hand column of the table. Excluded from the analysis were children whose dominant language was other than English (except for the Spanish-dominant children in California and Texas, who comprise their own analytic groups); also excluded were children with handicaps likely to impair test performance unduly.

Attrition Patterns

As can be seen in Table 2, the PDC testing sample has been reduced by 42% since fall 1976, the comparison testing sample by 29%. Overall, this represents an attrition rate of 36%. In the first phase of the PDC evaluation we had projected attrition rates of 37% for the PDC group, 42% for the comparison group, and 39% for the two groups combined. Table 3 shows that movement from the school area was the major reason for sample attrition. The differential attrition of PDC and comparison children is due to the fact that we continued to test comparison group children who moved from one local school to another as long as the new school did not house a PDC program. When PDC children moved, however, there were few or no other PDC schools to move to. Thus we did not continue testing these children, since they were no longer in a PDC program.

In some cases, children who are shown as having left the sample between fall 1976 and spring 1977 returned again in spring 1978. This occurred when children who left the Head Start program before 1977 testing were located in the "correct"

Table 1

Descriptive Characteristics of the Sample
Spring 1978

		Number in Full Sample	% Handicapped	ETHNICITY					SEX		LANGUAGE[a]		Number in Final Analytic Sample[b]
				% Black	% Hispanic	% American Indian/ Native Alaskan	% White	% Asian/Pacific Islander	% Female	% Male	% English	% Spanish	
CALIFORNIA-English	PDC	21	0	5	90	0	5	0	81	19	100	0	21
	Comp	17	6	6	59	0	35	0	65	35	100	0	16
CALIFORNIA-Spanish	PDC	4	0	0	100	0	0	0	50	50	0	100	4
	Comp	10	0	0	100	0	0	0	60	40	0	100	10
COLORADO	PDC	25	12	12	76	0	12	0	40	60	100	0	22
	Comp	22	9	0	64	0	36	0	50	50	100	0	20
CONNECTICUT	PDC	32	3	50	25	3	22	0	53	47	100	0	24
	Comp	46	2	83	11	0	7	0	41	59	100	0	44
FLORIDA	PDC	40	5	100	0	0	0	0	35	65	100	0	38
	Comp	34	0	91	6	0	3	0	47	53	100	0	34
GEORGIA	PDC	30	3	80	0	0	20	0	57	43	100	0	29
IOWA	PDC	25	4	44	4	0	52	0	52	48	100	0	24
	Comp	22	0	14	0	4	82	0	50	50	100	0	22
MARYLAND	PDC	32	25	47	9	0	41	3	44	56	100	0	24
	Comp	30	23	37	23	0	30	10	47	53	100	0	23
MICHIGAN	PDC	31	13	61	3	0	36	0	55	45	100	0	27
	Comp	51	8	75	2	0	24	0	55	45	100	0	47
TEXAS-English	PDC	29	0	3	55	0	41	0	45	55	100	0	18
	Comp	23	0	0	87	0	13	0	43	57	100	0	11
TEXAS-Spanish	PDC	17	0	0	100	0	0	0	35	65	0	100	17
	Comp	19	16	0	100	0	0	0	68	32	0	100	16
UTAH	PDC	28	18	0	21	0	79	0	46	54	100	0	23
	Comp	60	10	7	12	7	75	0	50	50	100	0	54
WASHINGTON	PDC	35	17	14	0	17	51	17	51	49	100	0	29
	Comp	47	13	45	2	2	49	2	60	40	100	0	41
WEST VIRGINIA	PDC	26	8	4	0	0	96	0	54	46	100	0	24
	Comp	26	19	12	0	0	88	0	69	31	100	0	21
TOTALS BY GROUP	PDC	375	9	36	25	2	35	2	49	51	94	6	324
	Comp	407	9	37	24	1	37	1	53	47	93	7	359
TOTALS, ALL GROUPS COMBINED		782	9	37	24	2	36	1	51	49	94	6	683

[a] In kindergarten year.

[b] Children who switched language groups from the Head Start to the kindergarten year were eliminated from the analytic sample, as were children with serious handicaps, based on Head Start year data.

Table 2

Summary of Year-to-Year Attrition[a]

	PDC						COMPARISON					
	Fall 1976 N	Spring 1977 N	Spring 1978 N	1976-1977 % Drop	1977-1978 % Drop	1976-1978 % Drop	Fall 1976 N	Spring 1977 N	Spring 1978 N	1976-1977 % Drop	1977-1978 % Drop	1976-1978 % Drop
California-English	37	31	21	16	32	43	25	21	17	16	19	32
California-Spanish	7	6	4	14	33	43	15	11	10	27	10	33
Colorado	55	50	25	9	50	55	32	25	22	22	12	31
Connecticut	56	53	32	5	40	43	57	54	46	5	17	19
Florida	47	39	40	17	--	15	39	35	14	10	3	13
Georgia	46	43	30	7	30	35	--- *No comparison sample.* ---					
Iowa	50	43	25	14	42	50	54	48	22	11	54	59
Maryland	44	41	32	7	22	27	58	47	30	19	36	48
Michigan	66	58	31	12	47	53	64	58	51	9	12	20
Texas-English	26	23	29[b]	12	--	--	20	19	23[b]	5	--	--
Texas-Spanish	38	31	17[b]	18	45	55	17	36	19[b]	3	47	49
Utah	68	64	28	6	56	59	61	54	60	12	--[c]	2
Washington	58	52	35	10	33	40	76	61	47	20	23	38
West Virginia	46	38	26	17	32	44	37	34	26	8	24	30
TOTAL	644	572	375	11	34	42	575	503	407	13	19	29

[a]Figures refer to number of children in the full sample, as distinguished from the reduced analytic sample.

[b]The language dominance classification (English/Spanish) of 24 children changed from the Head Start to the kindergarten year. Thus the composition of the English- and Spanish-dominant groups in California and Texas changed somewhat.

[c]Redesignation of group membership in Utah resulted in an increase in sample size from 1977 to 1978 (see text for explanation).

Table 3

Summary of Reasons for Attrition

Number of Children Not Tested Due To:

	Moving, Withdrawal		Continual Absence		Testing Refusal		Change in Group Designation		Other (including no permission slip)	
	1977	1978	1977	1978	1977	1978	1977	1978	1977	1978
PDC	63	176	2	0	2	0	0	27	4	16
COMPARISON	67	78	2	0	0	0	0	0	4	56
TOTAL	130	254	4	0	2	0	0	27	8	72

elementary school in 1978. There were 60 children in all who, in this way, left the sample and then returned. All of them had experienced at least three months of the Head Start program before departure.

Another situation affecting apparent attrition arose in Connecticut and Utah. In these sites, PDC and comparison children attended the same Head Start center and were designated "PDC" or "comparison" depending on the elementary school they were likely to attend. However, some of the "PDC" children went on to comparison schools and vice-versa. Since there had been no fundamental difference in the Head Start experiences of these two groups in either site, children who moved on to PDC and comparison schools were classified as members of those respective samples regardless of their original designation. These were the only instances in which children who switched groups were retained in the study.

A final note on attrition and its implications: although the attrition rate is somewhat lower than projected overall, among PDC children it is higher than we had estimated (42% instead of 37%). Added to this is the fact that an entire site, West Virginia, ceased participating in the program after the spring 1978 data were collected. Thus the number of PDC children likely to remain in the sample through the end of grade 3 is now almost certainly smaller than our earlier (March 1977) projection of 258. Our estimates at this point suggest a final PDC sample size of about 220--an average of about 20 PDC children per site. The size of the comparison sample at the end of grade 3 is likely to be larger than this, given current trends. In our judgment, these reduced samples will still be adequate for purposes of the evaluation.

Attrition's effects on group comparability. Attrition, beyond reducing sample size, can also have the effect of altering the group equivalence so important to long-term comparative studies like PDC. To examine this possibility, we conducted a series of analyses to assess the differences between the remaining PDC and comparison samples on major background characteristics and on performance measures obtained at the time of the children's entry into Head Start. The results of these analyses (chi-square tests for categorical variables, t tests for metric variables) are shown in Table 4 for the English-dominant analytic sample. On only two of the dimensions examined, POCL-1 and POCL-2, do the PDC and comparison groups differ significantly, and these differences are rather small (.14 standard deviations in the first case, .25 in the second). Furthermore, when all fall 1976 test scores are taken jointly as dependent variables in a multivariate

Table 4

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

(English-Dominant Analytic Sample)

	PDC Sample Spring 1978	Comparison Sample Spring 1978
N (approximate)	*303*	*333*
Background Characteristics		
Ethnicity (%)		
Black	42	42
Hispanic	17	15
American Indian/Native American	2	2
White	37	40
Asian/Pacific Islander	2	1
Sex (%)		
Male	47	48
Female	53	52
Prior Preschool (%)		
Yes	14	18
No	86	82
Age (months)	53.67	53.79
Number of Siblings	1.96	1.80
Mother's Education	11.02	11.13
Test Scores (Fall 1976)		
BSM-English	9.28	9.47
WPPSI	4.84	5.07
Verbal Fluency	6.41	6.41
Verbal Memory-1	12.99	13.68
Verbal Memory-3	2.60	2.76
Arm Coordination	3.43	3.61
Draw-A-Child	3.95	3.96
PIPS	2.03	2.15
POCL-1: "Task Orientation"	32.43	34.00*
POCL-2: "Sociability"	12.53	13.47*

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

Note: The PDC-comparison difference on all test scores taken jointly (MANOVA) is not statistically significant ($p > .10$).

analysis of variance, the difference between the profiles of the two groups does not approach statistical significance. And as Table 5 shows, the results of the same analyses for the Spanish-dominant analytic sample were equally positive (though it must be noted that the small size of the Spanish-dominant sample does not provide high power to detect group differences, should they exist).

All in all, the match that was accomplished between PDC and comparison groups in fall 1976 has withstood the effects of attrition, and group comparability remains excellent, at least for the English-dominant analytic sample. As for the Spanish-dominant analytic sample, our results suggest the same conclusion but we cannot state that conclusion with the same confidence, due to the low power of inference we command with that sample.

Attrition's effects on group representativeness. Attrition, whether it alters group comparability or not, can potentially affect the representativeness of the remaining groups. For example, if families of higher socioeconomic status were more mobile, then their children would be likelier to leave the PDC and comparison samples. If the trend were equal in both groups, this change in representation would not be revealed by analyses of the comparability of the present groups. Thus, to examine this additional contingency, we contrasted the children who have remained in the analytic samples through spring 1978 with those who have left, using the same dependent measures used in analyses of comparability. The results of these analyses are displayed in Tables 6 and 7. On the basis of the evidence given, we can conclude once again that attrition has not biased the remaining samples: for the PDC and comparison groups alike in the English- and Spanish-dominant samples, there are very few statistically significant differences on the dependent measures, taken one at a time. And when the dependent measures are examined jointly in multivariate analyses of variance, there are no indications whatever of significant differences between the profiles of the remaining and departing groups. This conclusion holds for all four series of analyses performed: on the English-dominant PDC group, the English-dominant comparison group, the Spanish-dominant PDC group, and the Spanish-dominant comparison group.

At this point, then, we can be confident in assuming that the remaining samples are representative of the original samples, established in fall 1976. Thus conclusions based on the experience of the present samples can be generalized to the Head Start population from which the PDC and comparison group children were drawn.

Table 5

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

(Spanish-Dominant Analytic Sample)

	PDC Sample *Spring 1978*	Comparison Sample *Spring 1978*
N (approximate)	21	26
Background Characteristics		
Ethnicity (%)		
Black	0	0
Hispanic	100	100
American Indian/Native American	0	0
White	0	0
Asian/Pacific Islander	0	0
Sex (%)		
Male	62	39
Female	38	62
Prior Preschool (%)		
Yes	14	35
No	86	65
Age (months)	54.81	55.65
Number of Siblings	2.38	3.08
Mother's Education	6.41	5.57
Test Scores (Fall 1976)		
BSM-English	.17	2.25
BSM-Spanish	12.75	12.50
WPPSI	8.71	7.00
Verbal Fluency	4.40	5.00
Verbal Memory-1	13.43	12.77
Verbal Memory-3	2.25	2.58
Arm Coordination	4.24	4.46
Draw-A-Child	4.25	4.69
PIPS	1.95	1.92
POCL-1: "Task Orientation"	30.10	31.35
POCL-2: "Sociability"	11.33	12.42

Note: The PDC-comparison difference on all variables singly and on all test scores jointly (MANOVA) is nonsignificant ($p > .10$).

Table 6

Representativeness of Remaining PDC and Comparison Children with Respect to Original (Fall 1976) Groups

(English-Dominant Analytic Sample)

	PDC		Comparison	
	Children Remaining in Testing Sample	Children Departing from Testing Sample	Children Remaining in Testing Sample	Children Departing from Testing Sample
N (approximate)	*303*	*205*	*333*	*160*
Background Characteristics				
Ethnicity (%)				
Black	42	35	42	34 *
Hispanic	17	22	15	24
American Indian/Native American	2	1	2	3
White	37	40	40	36
Asian/Pacific Islander	2	2	1	3
Sex (%)				
Male	47	53	48	47
Female	53	47	52	53
Prior Preschool (%)				
Yes	14	12	18	11*
No	86	88	82	89
Age (months)	53.68	53.43	53.79	53.68
Number of Siblings	1.96	1.77	1.80	1.99
Mother's Education	11.02	11.15	11.13	11.14
Test Scores (Fall 1976)				
BSM-English	9.28	9.40	9.47	9.58
WPPSI	4.84	5.11	5.07	5.11
Verbal Fluency	6.41	5.39*	6.41	6.25

*Difference on this variable between remaining and departing groups significant with $p < .10$ (two-tailed).

Table 6
(continued)

	PDC		Comparison	
	Children Remaining in Testing Sample	Children Departing from Testing Sample	Children Remaining in Testing Sample	Children Departing from Testing Sample
Verbal Memory-1	12.99	12.19	13.68	13.90
Verbal Memory-3	2.60	2.70	2.76	2.54
Arm Coordination	3.43	3.46	3.61	3.14
Draw-A-Child	3.95	3.88	3.96	4.30
PIPS	2.03	1.89	2.15	2.05
POCL-1: "Task Orientation"	32.43	32.89	34.00	32.35*
POCL-2: "Sociability"	12.53	12.76	13.47	12.46*

*Difference on this variable between remaining and departing groups significant with $p < .10$ (two-tailed).

Note: The difference between remaining and departing children on all test scores taken jointly (MANOVA) is nonsignificant ($p > .10$). This is true for both the PDC and comparison groups.

Table 7

Representativeness of Remaining PDC and Comparison Children with Respect to Original (Fall 1976) Groups
(Spanish-Dominant Analytic Sample)

	PDC		Comparison	
	Children Remaining in Testing Sample	Children Departing from Testing Sample	Children Remaining in Testing Sample	Children Departing from Testing Sample
N (approximate)	*21*	*13*	*26*	*11*
Background Characteristics				
Ethnicity (%)				
Black	0	0	0	0
Hispanic	100	100	100	100
American Indian/Native American	0	0	0	0
White	0	0	0	0
Asian/Pacific Islander	0	0	0	0
Sex (%)				
Male	62	50	39	36
Female	38	50	61	64
Prior Preschool (%)				
Yes	14	15	35	55
No	86	85	65	45
Age (months)	54.81	55.08	55.65	55.82
Number of Siblings	2.38	1.33*	3.08	2.18
Mother's Education	6.41	7.11	5.57	6.55
Test Scores (Fall 1976)				
BSM-English	.17	3.00*	2.25	4.83
BSM-Spanish	12.75	13.18	12.50	10.36
WPPSI	8.71	7.23	7.00	5.55
Verbal Fluency	4.40	3.00	5.00	5.64
Verbal Memory-1	13.43	13.00	12.77	13.00
Verbal Memory-3	2.25	2.23	2.58	1.82

*Difference on this variable between remaining and departing groups significant with $p < .10$ (two-tailed).

Table 7

(continued)

	PDC		Comparison	
	Children Remaining in Testing Sample	Children Departing from Testing Sample	Children Remaining in Testing Sample	Children Departing from Testing Sample
Arm Coordination	4.24	4.13	4.46	3.03
Draw-A-Child	4.25	4.25	4.69	4.45
PIPS	1.95	1.67	1.92	2.09
POCL-1: "Task Orientation"	30.10	26.62	31.35	30.18
POCL-2: "Sociability"	11.33	10.15	12.42	12.55

*Difference on this variable between remaining and departing groups significant with $p < .10$ (two-tailed).

Note: The difference between remaining and departing children on all test scores taken jointly (MANOVA) is nonsignificant ($p > .10$). This is true for both the PDC and comparison groups.

Characteristics of the Child Measures

Review of Past Findings

One of the major tasks of Phase 1 of the PDC evaluation was to examine the adequacy of each of the instruments in the PDC battery. Criteria were set regarding critical measurement properties, and the suitability of each instrument was judged against its showing on these criteria. On this basis, a number of instruments were deleted from the battery or revised. The instruments that were retained constituted a battery that, in our judgment, satisfied most or all of the conditions that could reasonably be applied. A summary of past psychometric findings is given in Table 8. Descriptions of the instruments themselves appear in Appendix A.

Revision of the PDC Child Rating Scale

The Child Rating Scale (CRS), as administered in 1976, 1977, and 1978, has consisted of 39 items, all revolving around the child's behavior as observed by the classroom teacher. Before proceeding with this year's analysis of CRS data, we undertook to determine whether it would be possible in the future to reduce the number of items while retaining the integrity of the scale, thereby reducing the rating burden on the teacher without diminishing the measure's utility to the evaluation. The analyses aimed at reducing the CRS proceeded in several steps, described below.

1. In the past, factor analysis of the 39-item CRS has consistently yielded six fairly stable factors. Analyses of PDC-comparison differences on the CRS have been based upon scale scores formed by summing each child's ratings on the items associated with each of the six factors. The first step in our reduction procedure was to determine which of the items in each of these scales were the best predictors of the total scale score. We began with CRS data from spring 1977, allowing for independent replication of the results later with spring 1978 data. Using regression techniques, we selected the smallest set of items that predicted each of the six factor scores with at least 90% precision ($R^2 > .90$). These sets, which turned out to consist of only 17 items in all, represented reduced versions of the six empirically determined scales.

Table 8

Summary of the Psychometric Characteristics of the Battery in Fall 1976 and Spring 1977
(Based on Data for Children in the Analytic Sample)

MEASURE	Internal Consistency (Cronbach's Alpha)		Validity: Acceptable evidence?	Stability (Fall-spring correlation)	Sensitivity to change: Was actual spring mean > than predicted spring mean?	Factor Structure[e]		Relationship to social competence: % of variance accounted for by "social competence" criteria beyond background varibles (s77)
	Fall 1976	Spring 1977				Fall 1976	Spring 1977	
ENGLISH-DOMINANT CHILDREN								
BSM-English	.84	.83	yes	.71	no	2	1	.07
BSM-Spanish	.95	.94	d	d	d	d	d	d
Verbal Fluency	.76	.76	yes	.49	yes	1	1	.19
Verbal Memory-1	.85	.82	yes	.51	yes	1	1	.13
Verbal Memory-3	.82	.81	yes	.47	yes	1	1	.20
Draw-A-Child	.84	.81	yes	.54	yes	2	1	.11
Arm Coordination	.65	.69	yes	.43	yes	3	4	.03
PIPS	a	a	yes	.38	d	1	1	.12
POCL-1: "Task Orientation"	.95	.95	d	.44	d	1	3	
POCL-2: "Sociability"	.90	.87	d	.43	d	1	3	
CRS-1: "Friendliness"	c	.85	d	c	d	d	2	
CRS-2: "Aggressiveness"	c	.84	d	c	d	d	3	
CRS-3: "Perseverance"	c	.89	d	c	d	d	2	
CRS-4: Independence"	c	.77	d	c	d	d	4	
CRS-5: "Self-Assurance"	c	.91	d	c	d	d	2	
CRS-6: "Resourcefulness"	c	.74	d	c	d	d	2	
SPANISH-DOMINANT CHILDREN								
BSM-English	.93	.90	d	.90	d	f	f	
BSM-Spanish	.86	.70	yes	.59	d	2	1	f
Verbal Fluency	.81	.76	yes	.52	d	1	1	f
Verbal Memory-1	.89	.85	yes	.69	d	1	1	f
Verbal Memory-3	.84	.78	yes	.39	d	1	1	f
Draw-A-Child	.78	.69	yes	.39	d	1	1	f
Arm Coordination	.73	.71	yes	.24	d	3	2	f
PIPS	a	a	yes	.17	d	2	1	f
POCL-1: "Task Orientation"	.96	.94	yes	.39	d	2	2	f
POCL-2: "Sociability"	.96	.90	yes	.39	d	2	2	f

[a]Consists of a single composite item, thus internal consistency coefficient cannot be computed.

[b]The subscales of the Child Rating Scale have been reconstituted since the analyses summarized here were done.

[c]Not administered in the fall.

[d]Not analyzed.

[e]These columns show which measures loaded on each of the factors (1, 2, 3, etc.) emerging from the various factor analyses.

[f]Small sample size did not permit this analysis for the Spanish-dominant sample.

2. We then factor analyzed these 17 items to determine whether the same six-factor structure would emerge. Only three factors were actually found, but they bore a logical relationship to their "ancestor" factors that seemed to support the assumption that they adequately represented the original structure (just as factor analysis reduces a complex data structure to a simpler pattern).

3. We calculated internal consistency (Cronbach's alpha) coefficients for each reduced scale and found that all coefficients exceeded .75, indicating adequate homogeneity of the scales' constituent items.

4. Using scores based on the six scales of the full 39-item CRS we performed regression analyses to determine the relationship between these six scores and children's scores on all other tests in the PDC battery. (These are the analyses we have presented in the past under the rubric "relationship to social competence.") Next, we performed parallel analyses using the three scale scores yielded by the reduced version of the CRS. We then compared the coefficients of determination (R^2 values) obtained in these parallel analyses to learn how much predictive precision would be lost if we relied in the future on the reduced 17-item version rather than the original 39-item version. Typically, precision was reduced by a margin of only 1 or 2% (e.g., from an R^2 value of .19 to a value of .18 or .17).

5. Finally, turning to spring 1978 data, we assembled scores on the three reduced scales based on the 17-item version. Factor analyses of these items replicated the factor structure found in spring 1977 data. Internal consistency coefficients based on 1978 data also paralleled those of the preceding year.

Having arrived at a version of the CRS that reduced the number of constituent items by more than half while sacrificing only slight predictive power, we decided to use the reduced version as the basis for analysis in this and future reports. Thus for this report we have constructed 1977 and 1978 CRS scores for each child that are based only on the 17 items that comprise the reduced version of the CRS, even though teachers completed the 39-item version in those years. From spring 1979 through spring 1981, the CRS form that teachers complete will consist of the 17 items referred to here plus three more, hypothetically related to an "academic motivation" factor. These were added to provide coverage of a dimension of behavior that was not explicitly represented on the original CRS. All 20 items of the new CRS are presented in Table 9.

Table 9

Item Clusters Produced by Factor Analyses of the Child Rating Scale[a]

Item No. on Revised 17-item CRS	
	CRS-1: "Self-Assurance"
1	Shows self-confidence
5	Shows respect for or tolerance of others' ideas and behavior or looks
7	Is easily distracted when doing a task
8	Attempts to solve social problems with little adult assistance
10	Gets the attention of peers appropriately
11	Returns to unfinished tasks after interruption
14	Cooperates and shares with others
15	Recognizes others' feelings, responds appropriately
16	Enjoys tasks he/she chooses
18	Has a desire to master all kinds of skills
19	Talks freely to children
	CRS-2: "Aggressiveness"
2	Shows verbal dislike or hostility to others
13	Uses words or wits to try to influence others
17	Uses physical force to try to control others
20	Competes with others for toys, attention, achievement
	CRS-3: "Dependence"
3	Is controlled or influenced by others
6	Imitates others or follows them around

[a]Three more items have been added to those that emerged from empirical analysis of the original CRS, to represent an "academic motivation" factor:

4. Is motivated toward academic performance
9. Completes assignments
12. Is alert and interested in school work

Psychometric Analyses of Spring 1978 Data

Score distributions. Table 10 gives the means and standard deviations of each measure used in the 1978 battery. These statistics are based upon data drawn from the samples of PDC and comparison children who have remained in the evaluation through spring 1978. (Some of these children were not tested on some measures in fall 1976 or spring 1977, even though they were present, thus the sample sizes for those years tend to be smaller than they are for 1978.)

The year-to-year status of PDC and comparison children on these measures is represented graphically in Figures 1a to 1g (for the English-dominant sample) and 2a to 2g (for the Spanish-dominant sample). To allow comparison from measure to measure and from year to year, each of the group means graphed has been converted to a z score that is based upon the overall mean and standard deviation for that measure in its first administration (fall 1976 for most of the measures; spring 1977 for the Child Rating Scale scores; the PIAT had never been administered before spring 1978).[1]

The graphs presented in Figures 1 and 2 allow examination of the relative overall status of PDC and comparison children at any timepoint on any measures in the battery. The graphs also depict changes over time in level of performance relative to earlier performance. Thus an "uphill" curve indicates

---

[1]The means plotted are standardized according to the formula:

$$z_{ij} = \frac{\bar{X}_{ij} - \bar{X}'}{S'}$$

where

- i = treatment group (PDC/comparison),
- j = assessment period (fall 1976, spring 1977, spring 1978),
- z = standardized mean,
- $\bar{X}$ = raw score mean,
- $\bar{X}'$ = raw score mean for PDC and comparison groups, combined, in baseline year, and
- S' = standard deviation for PDC and comparison groups, combined, in baseline year.

Table 10

Means and Standard Deviations of Child Measures for Three Timepoints: Fall 1976, Spring 1977 and Spring 1978 (Based on Children Still Present in 1978)

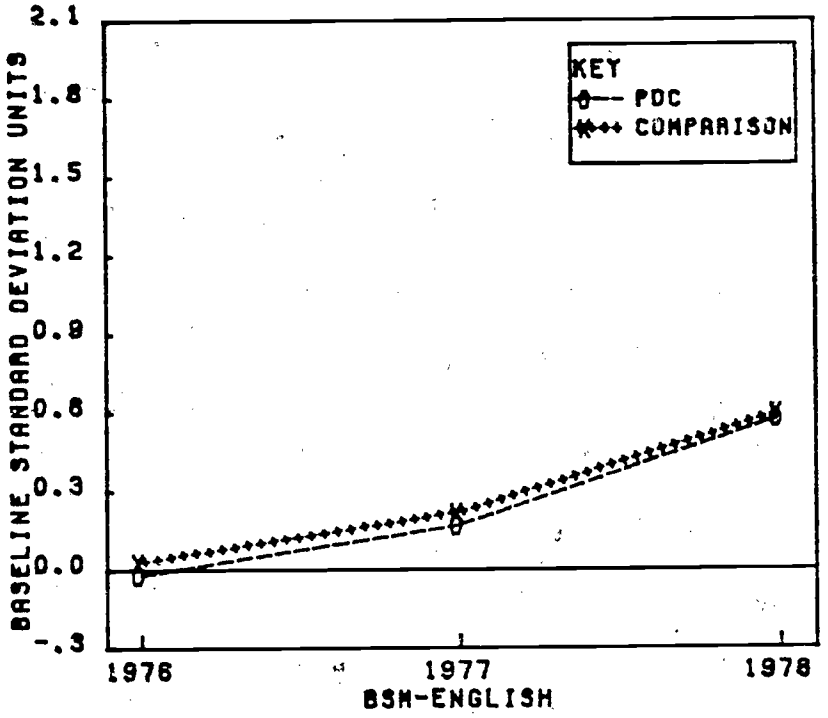
Measures	Fall 1976			Spring 1977			Spring 1978		
	N	Mean	SD	*N*	Mean	SD	*N*	Mean	SD
ENGLISH-DOMINANT CHILDREN									
BSM-English	635	9.38	4.27	595	10.26	4.01	636	11.93	3.09
BSM-Spanish	15	6.40	5.97	9	4.78	4.29	19	7.26	6.26
Verbal Fluency	626	6.41	4.98	596	9.34	5.73	636	14.57	5.60
Verbal Memory-1	635	13.35	6.47	597	14.83	6.57	636	18.22	5.31
Verbal Memory-3	631	2.68	2.55	593	3.87	2.74	636	5.75	2.54
Arm Coordination	625	3.53	3.04	596	4.57	3.35	636	8.11	4.66
Draw-A-Child	626	3.96	2.85	597	5.69	2.67	636	7.63	1.80
PIPS	627	2.09	1.65	596	2.66	1.77	636	3.75	1.69
PIAT-Math[a]							615	13.40	3.96
PIAT-Reading[a]							600	15.42	4.30
POCL-1: "Task Orientation"	636	33.26	9.40	565	34.37	9.74	595	36.84	8.57
POCL-2: "Sociability"	636	13.02	3.87	565	13.46	3.82	595	13.42	3.68
CRS-1: "Self-Assurance"[b]				524	35.89	6.95	553	35.18	7.71
CRS-2: "Aggressiveness"[b]				544	11.62	3.03	574	11.53	3.56
CRS-3: "Dependence"[b]				546	5.72	1.59	572	5.46	1.84
SPANISH-DOMINANT CHILDREN									
BSM-English	14	1.36	3.41	7	4.57	3.74	36	8.50	4.13
BSM-Spanish	46	12.61	4.12	43	12.81	2.33	46	13.13	2.45
Verbal Fluency	46	4.74	4.37	45	7.42	4.35	47	9.47	3.63
Verbal Memory-1	47	13.06	7.34	45	16.38	6.66	47	18.26	6.57
Verbal Memory-3	46	2.43	2.46	45	4.18	2.67	46	5.11	2.78
Arm Coordination	46	4.37	3.83	45	5.94	3.70	47	9.25	4.74
Draw-A-Child	46	4.50	2.28	45	6.44	1.97	47	7.70	2.04
PIAT-Math[a]							26	11.92	3.04
PIAT-Reading[a]							26	10.54	4.41
PIPS	46	1.93	1.31	45	3.07	1.89	47	4.00	1.86
POCL-1: "Task Orientation"	47	30.79	10.15	45	37.53	8.71	42	35.93	8.43
POCL-2: "Sociability"	47	11.94	3.82	45	13.40	4.05	42	13.17	3.18
CRS-1: "Self-Assurance"[b]				4	38.25	4.79	41	34.73	5.63
CRS-2: "Aggressiveness"[b]				4	10.25	2.36	46	10.94	2.74
CRS-3: "Dependence"[b]				4	5.25	0.96	45	6.04	1.59

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

[b]The Child Rating Scale is only administered in the spring, thus there are no fall 1976 figures.

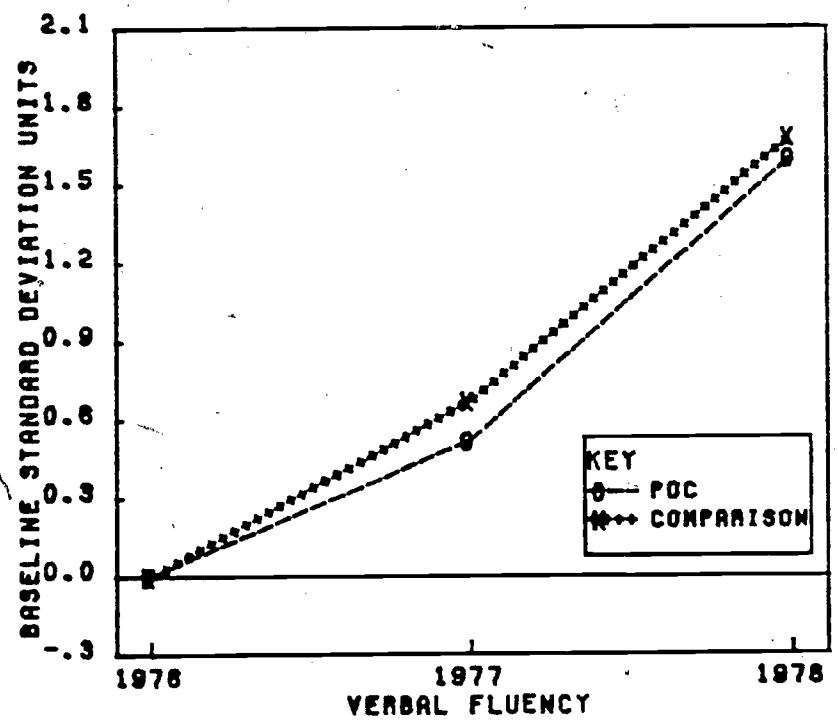
Figure 1a

Plot of PDC and Comparison Group Means on Measures in the Fall 1976, Spring 1977, and Spring 1978 Batteries (English-Dominant Analytic Sample)



N (PDC) =281

N (COMPARISON) =314



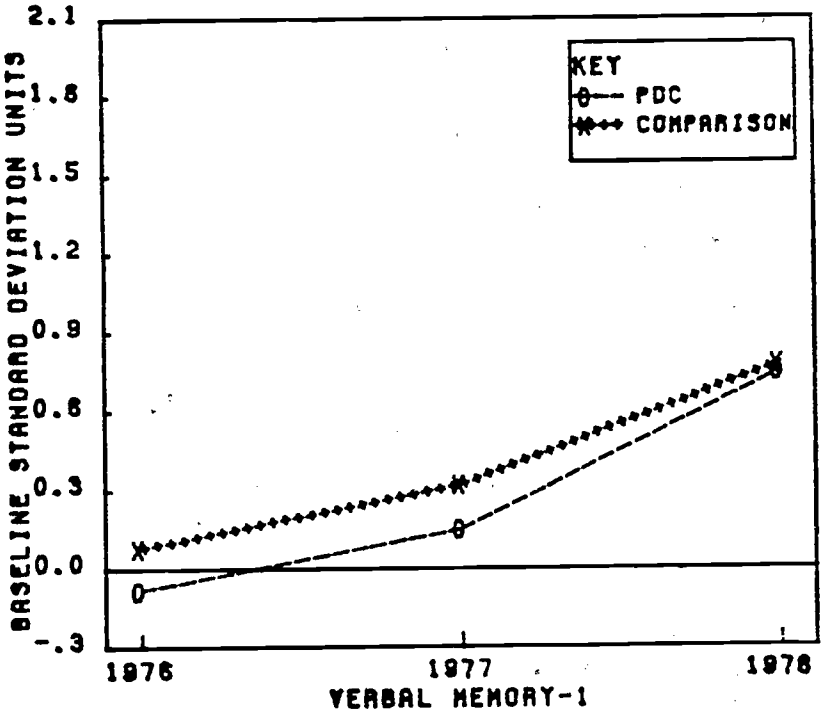
N (PDC) =277

N (COMPARISON) =309

Note: Each of the means plotted is standardized with reference to fall 1976 scores on the same measure (see text for technical detail and explanation of exceptions). The means for each variable are based on cases with complete data for all timepoints shown.

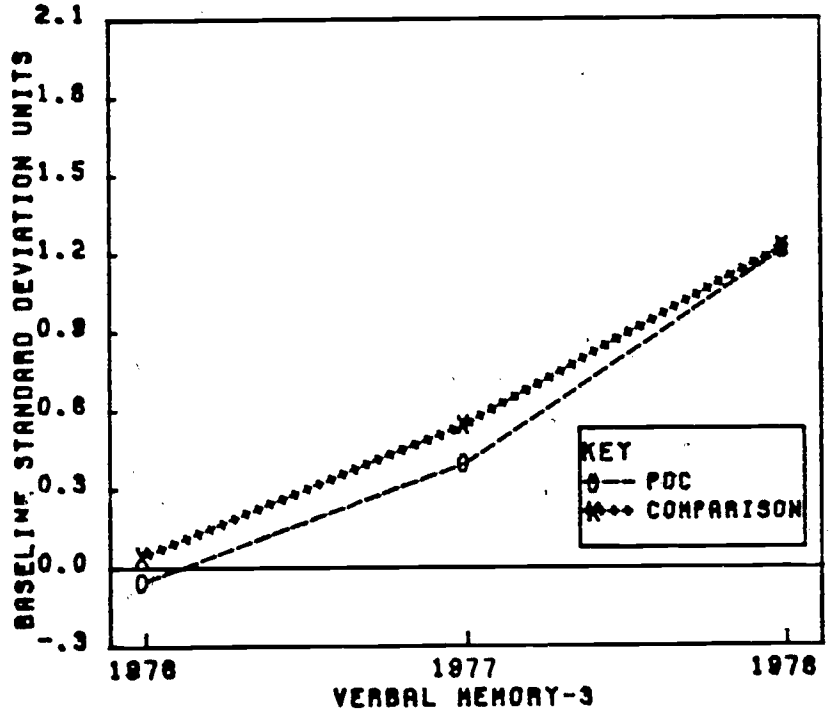
Figure 1b

Plot of PDC and Comparison Group Means on Measures in the Fall 1976, Spring 1977, and Spring 1978 Batteries (English-Dominant Analytic Sample)



N (PDC) =283

N (COMPARISON) =313



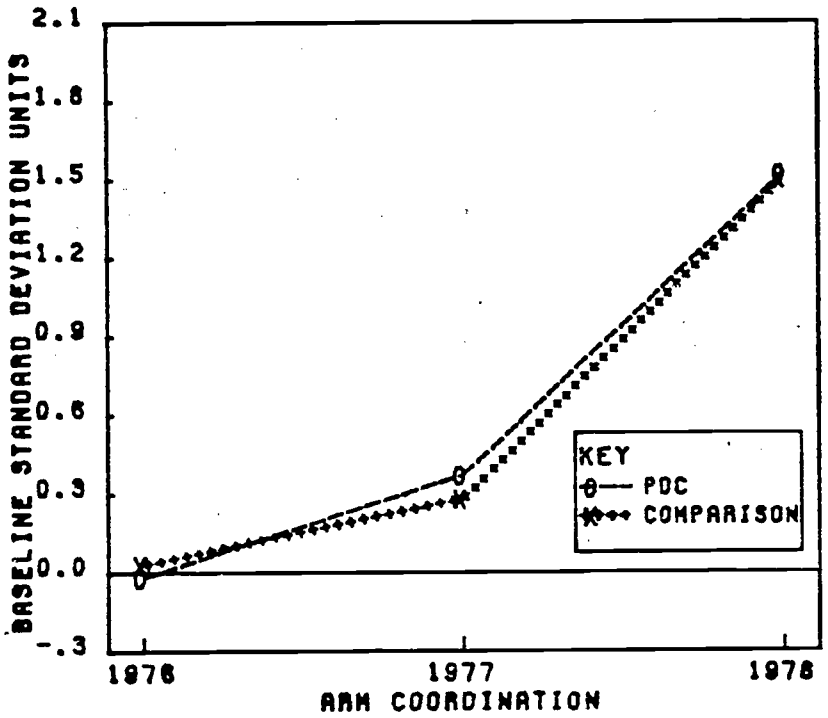
N (PDC) =279

N (COMPARISON) =309

Note: Each of the means plotted is standardized with reference to fall 1976 scores on the same measure (see text for technical detail and explanation of exceptions). The means for each variable are based on cases with complete data for all timepoints shown.

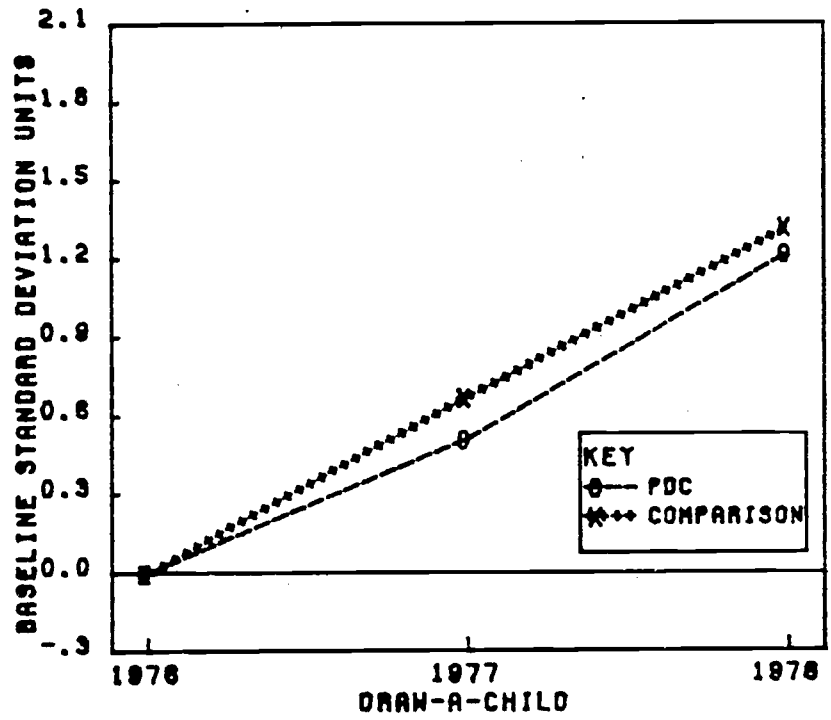
Figure 1c

Plot of PDC and Comparison Group Means on Measures in the Fall 1976, Spring 1977, and Spring 1978 Batteries (English-Dominant Analytic Sample)



N (PDC) =277

N (COMPARISON) =308



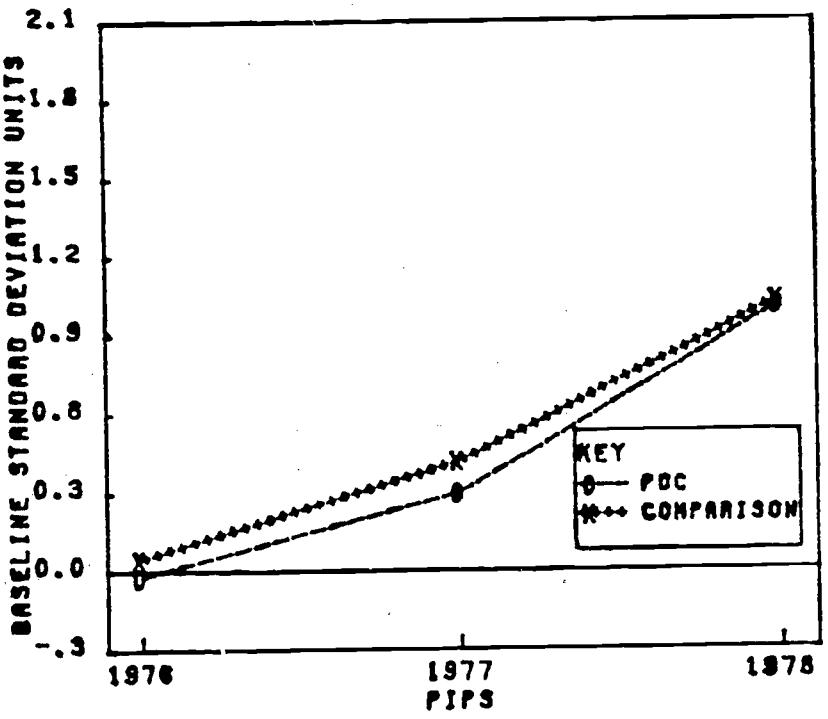
N (PDC) =278

N (COMPARISON) =309

Note: Each of the means plotted is standardized with reference to fall 1976 scores on the same measure (see text for technical detail and explanation of exceptions). The means for each variable are based on cases with complete data for all timepoints shown.

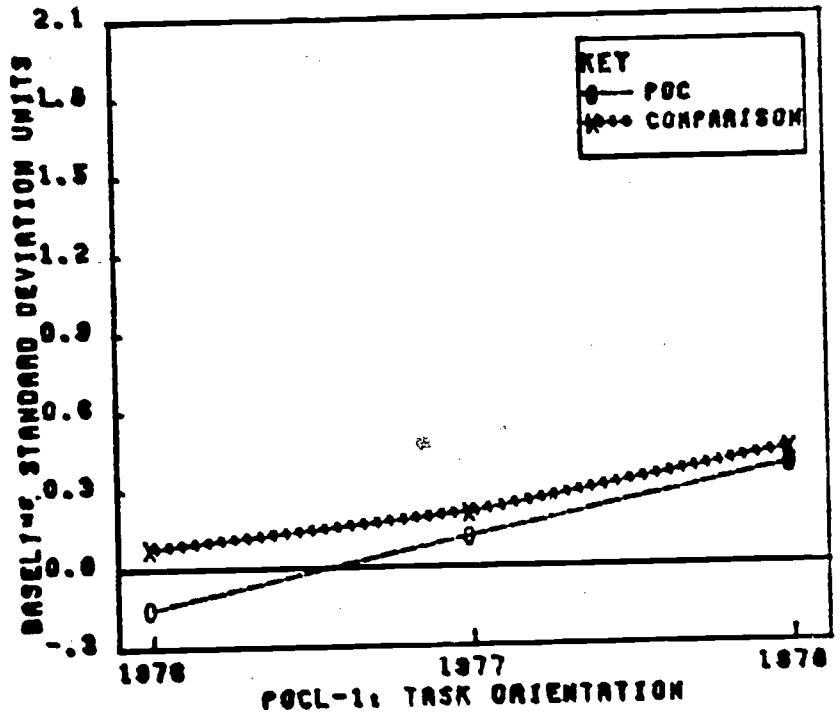
Figure 1d

Plot of PDC and Comparison Group Means on Measures in the Fall 1976, Spring 1977, and Spring 1978 Batteries (English-Dominant Analytic Sample)



N (PDC) =278

N (COMPARISON) =309



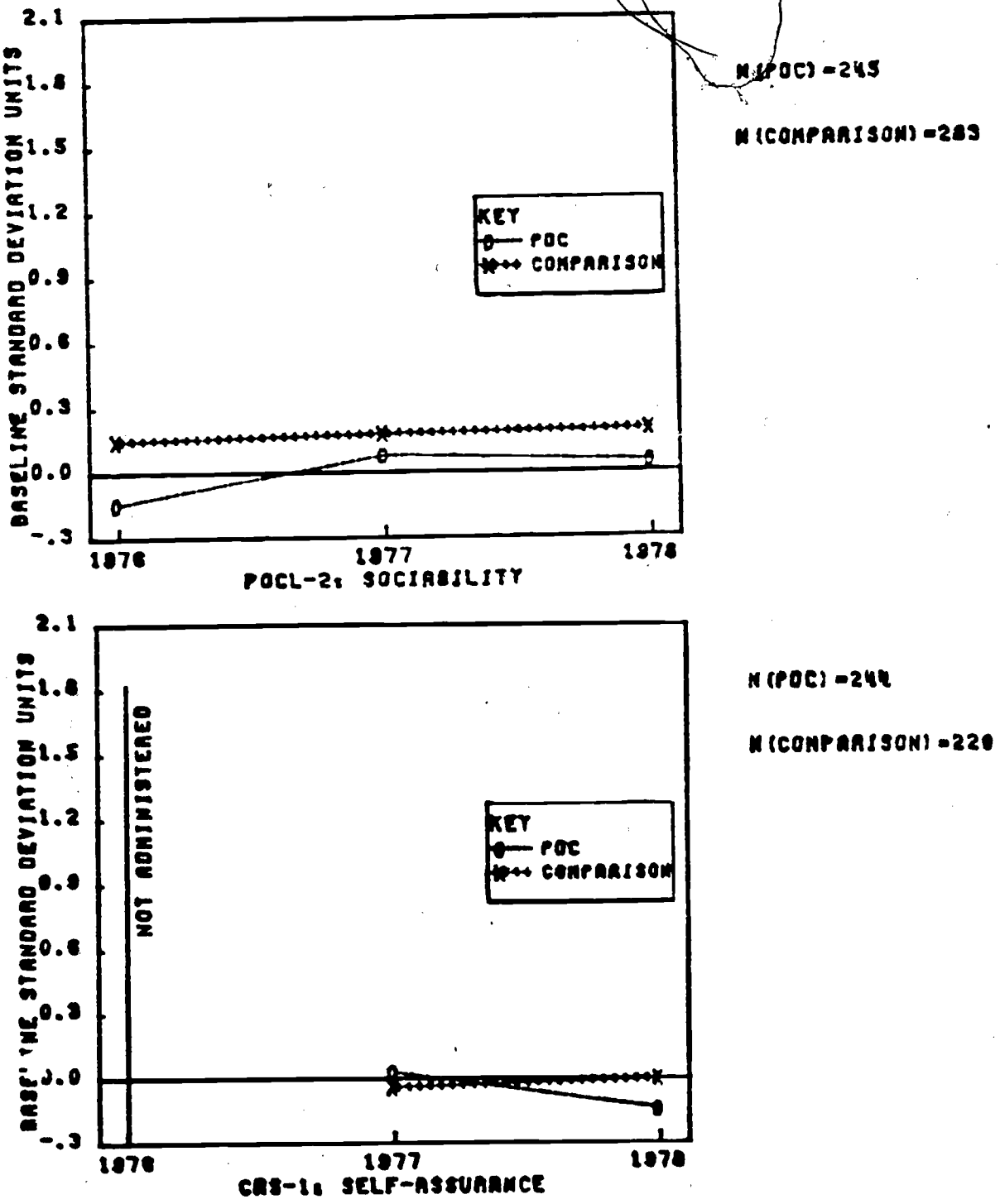
N (PDC) =245

N (COMPARISON) =283

Note: Each of the means plotted is standardized with reference to fall 1976 scores on the same measure (see text for technical detail and explanation of exceptions). The means for each variable are based on cases with complete data for all timepoints shown.

Figure 1e

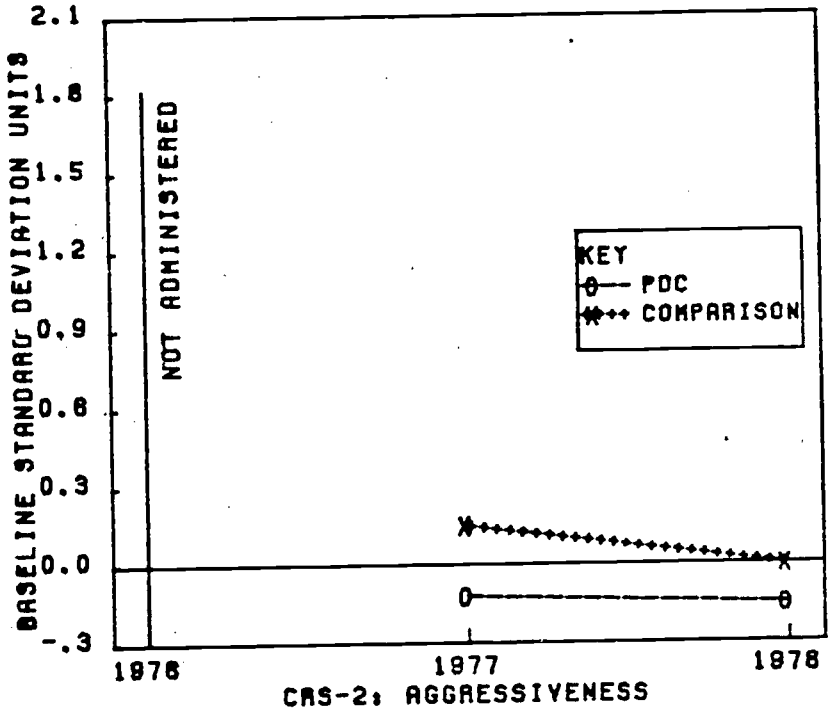
Plot of PDC and Comparison Group Means on Measures in the Fall 1976, Spring 1977, and Spring 1978 Batteries (English-Dominant Analytic Sample)



Note: Each of the means plotted is standardized with reference to fall 1976 scores on the same measure (see text for technical detail and explanation of exceptions). The means for each variable are based on cases with complete data for all timepoints shown.

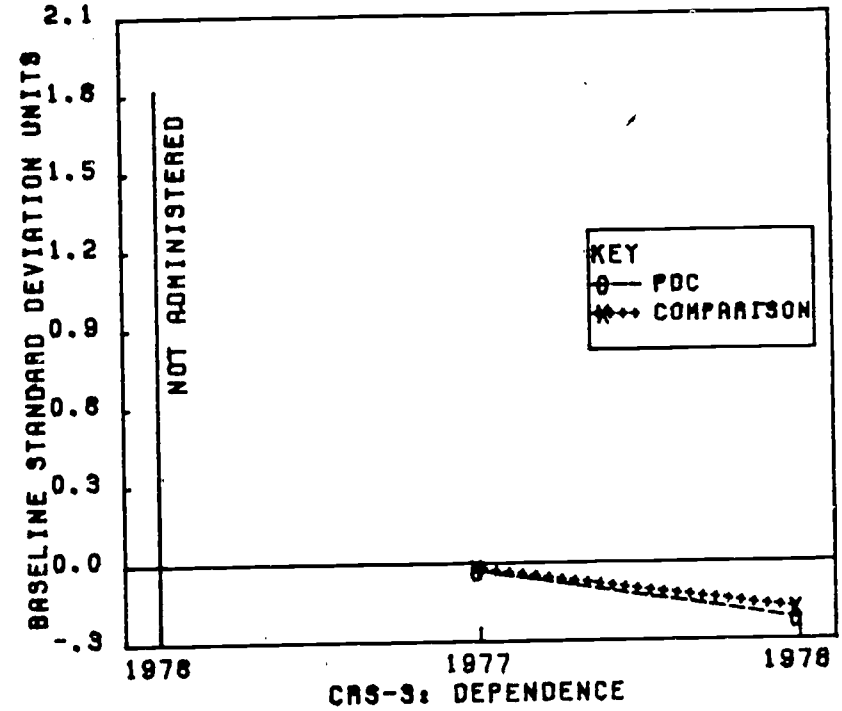
Figure 1f

Plot of PDC and Comparison Group Means on Measures in the Fall 1976, Spring 1977, and Spring 1978 Batteries (English-Dominant Analytic Sample)



N (PDC) =260

N (COMPARISON) =235



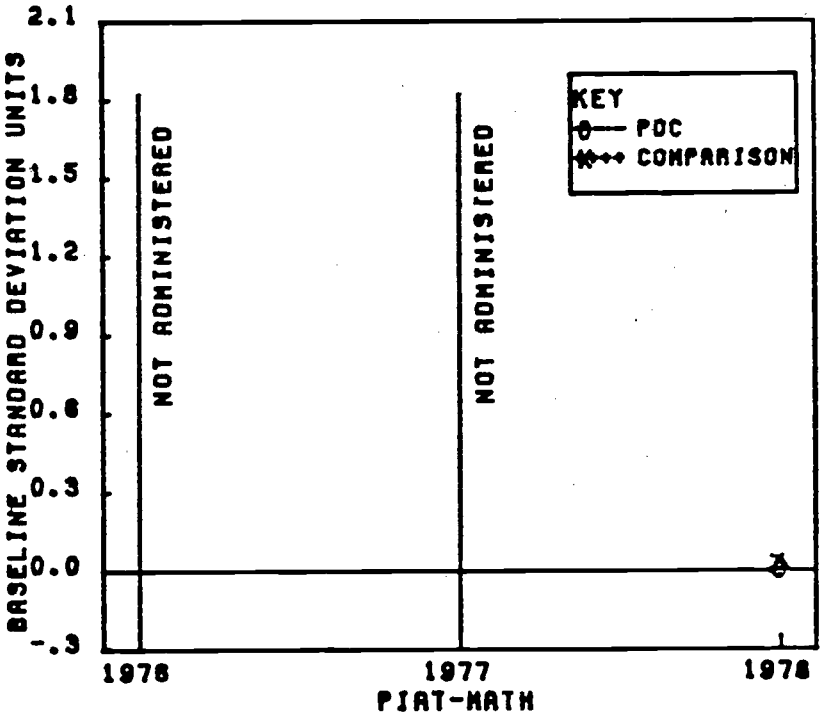
N (PDC) =261

N (COMPARISON) =236

Note: Each of the means plotted is standardized with reference to fall 1976 scores on the same measure (see text for technical detail and explanation of exceptions). The means for each variable are based on cases with complete data for all timepoints shown.

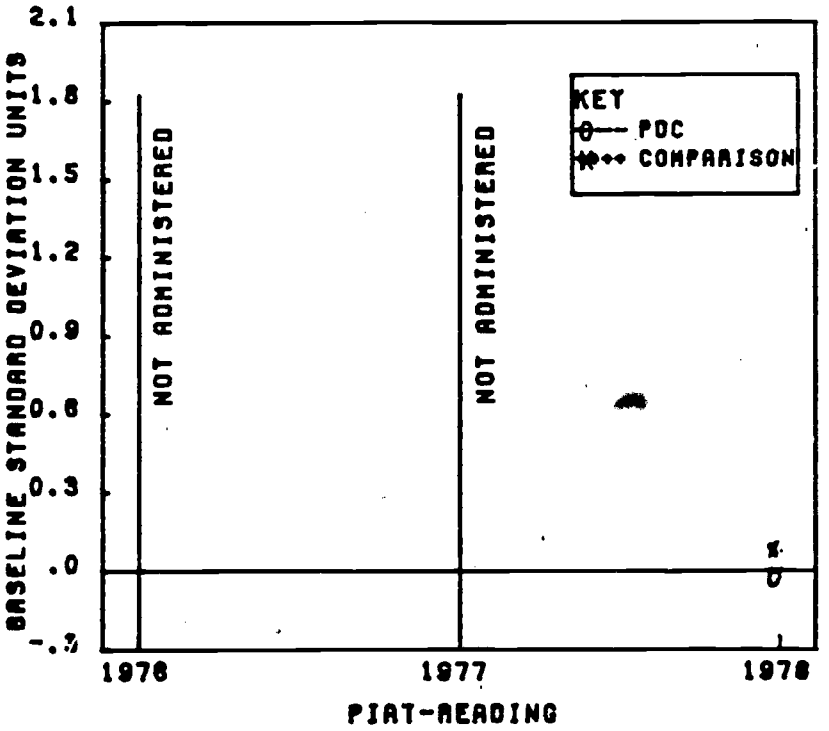
Figure 1g

Plot of PDC and Comparison Group Means on Measures in the Fall 1976, Spring 1977, and Spring 1978 Batteries (English-Dominant Analytic Sample)



N (PDC) =295

N (COMPARISON) =320



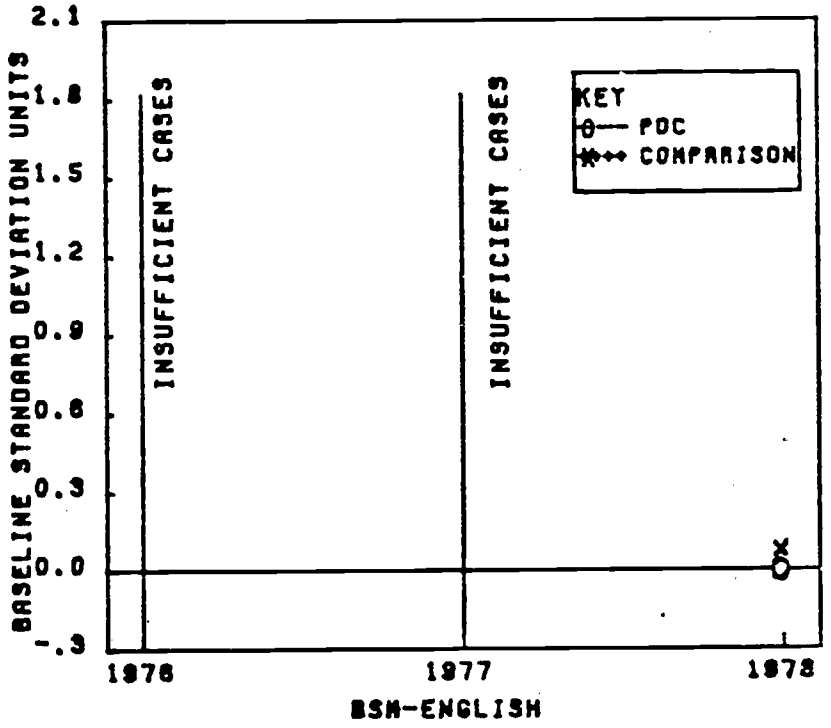
N (PDC) =286

N (COMPARISON) =314

Note: Each of the means plotted is standardized with reference to fall 1976 scores on the same measure (see text for technical detail and explanation of exceptions). The means for each variable are based on cases with complete data for all timepoints shown.

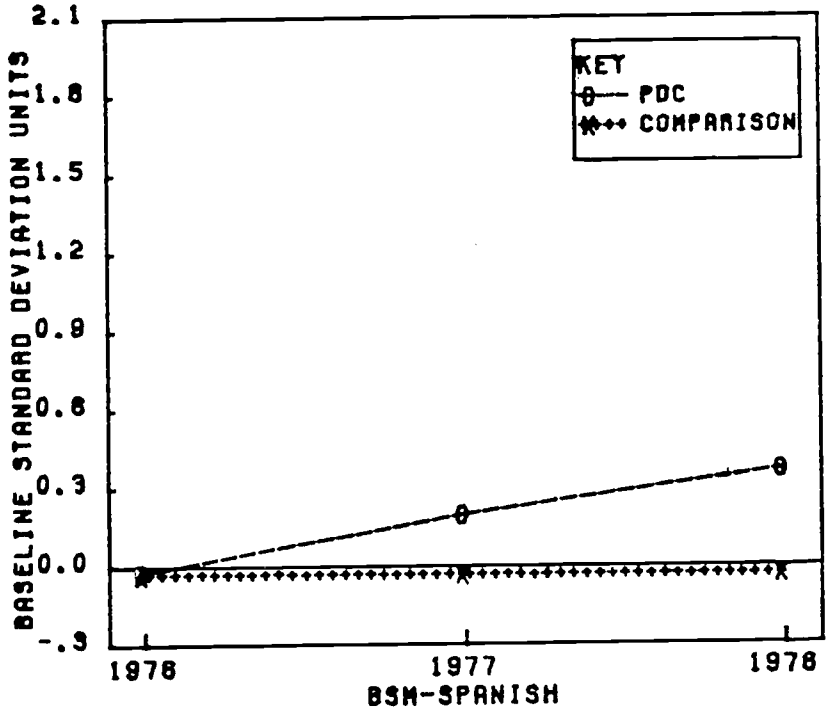
Figure 2 a

Plot of PDC and Comparison Group Means on Measures in the Fall 1976, Spring 1977, and Spring 1978 Batteries (Spanish-Dominant Analytic Sample)



N (PDC) = 11

N (COMPARISON) = 22



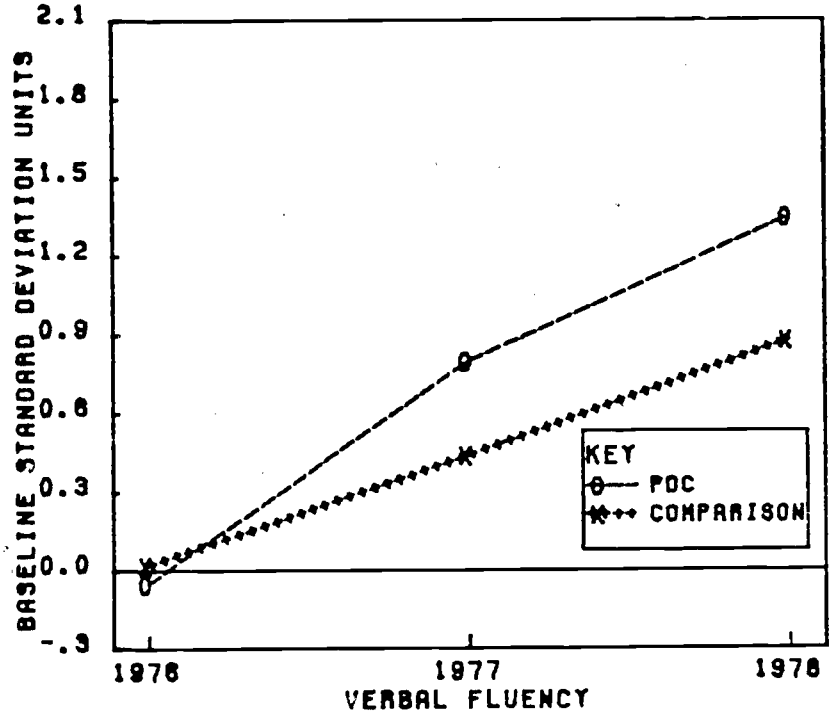
N (PDC) = 16

N (COMPARISON) = 25

Note: Each of the means plotted is standardized with reference to fall 1976 scores on the same measure (see text for technical detail and explanation of exceptions). The means for each variable are based on cases with complete data for all timepoints shown.

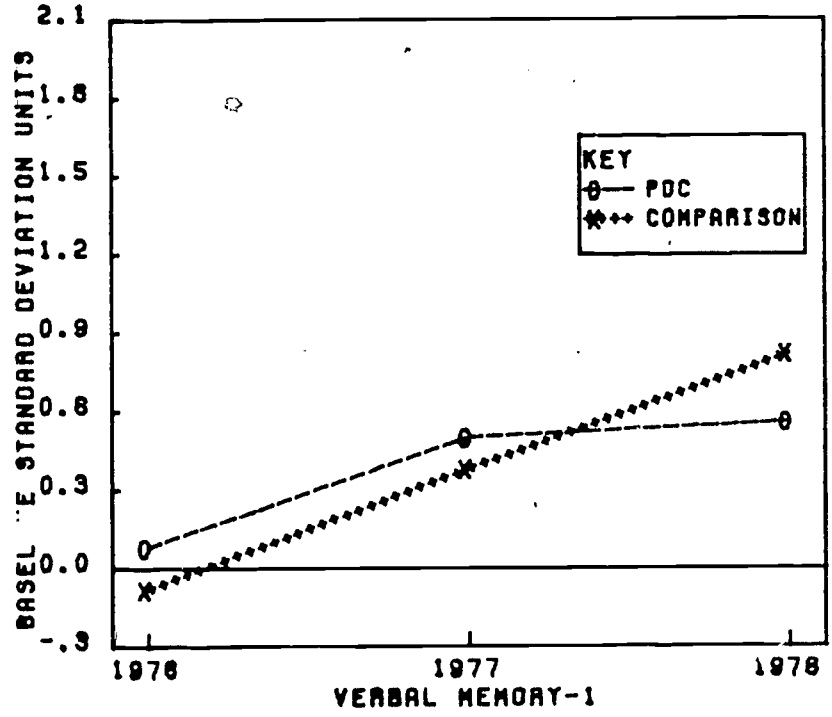
Figure 2 b

Plot of PDC and Comparison Group Means on Measures in the Fall 1976, Spring 1977, and Spring 1978 Batteries (Spanish-Dominant Analytic Sample)



N(PDC) =19

N(COMPARISON) =26



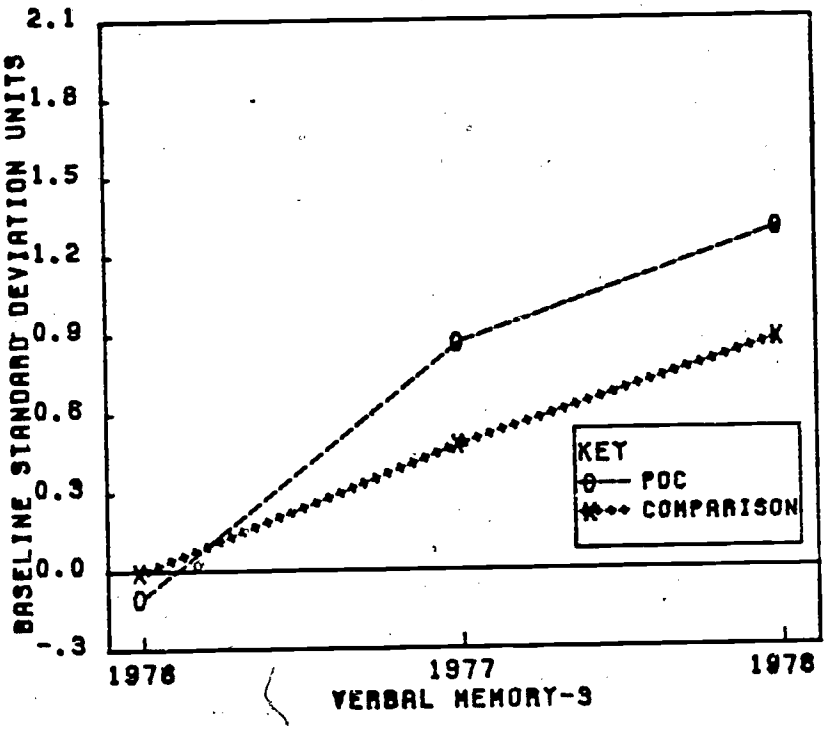
N(PDC) =19

N(COMPARISON) =26

Note: Each of the means plotted is standardized with reference to fall 1976 scores on the same measure (see text for technical detail and explanation of exceptions). The means for each variable are based on cases with complete data for all timepoints shown.

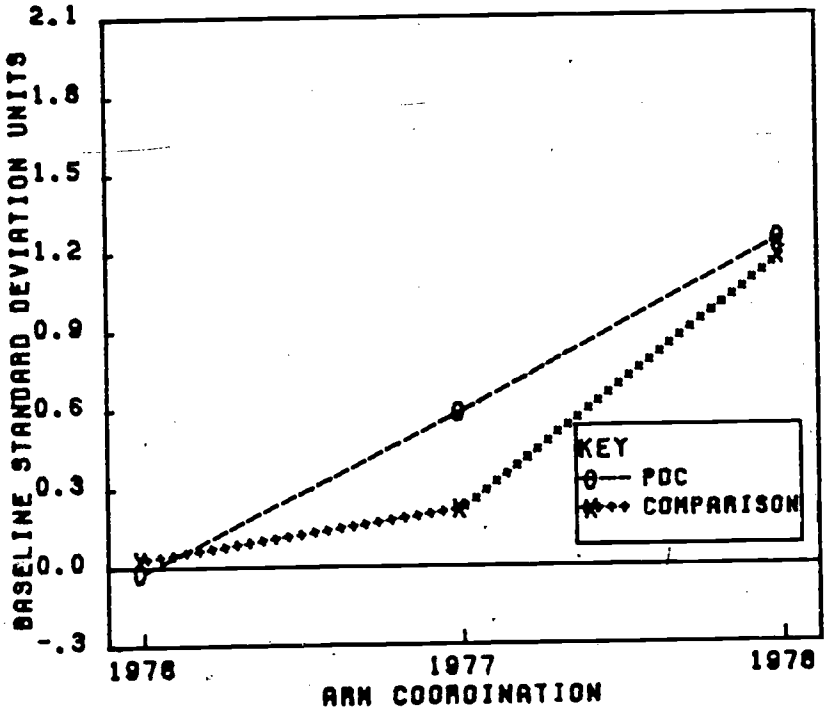
Figure 2c

Plot of PDC and Comparison Group Means on Measures in the Fall 1976, Spring 1977, and Spring 1978 Batteries (Spanish-Dominant Analytic Sample)



N (PDC) =17

N (COMPARISON) =26



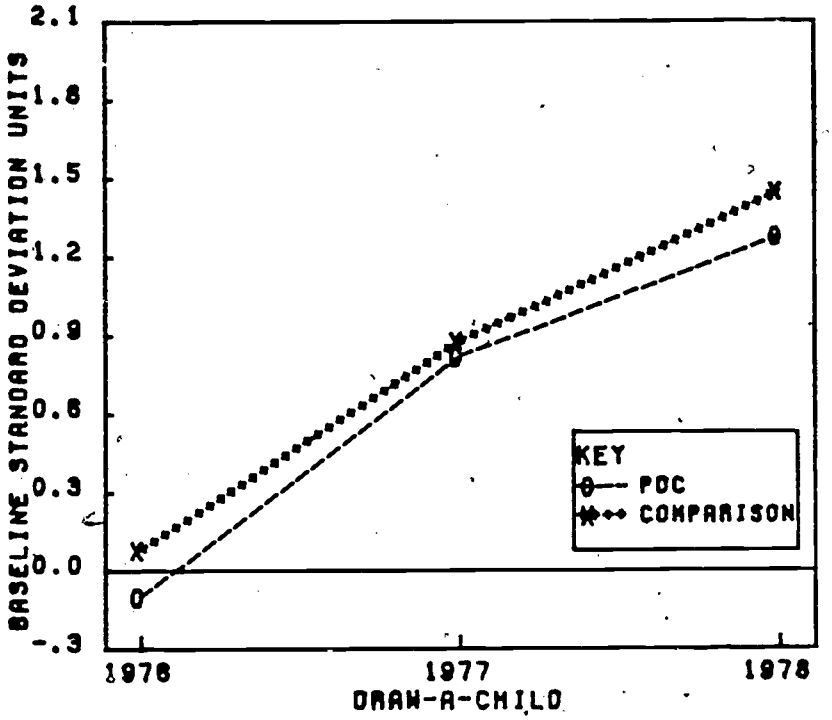
N (PDC) =19

N (COMPARISON) =26

Note: Each of the means plotted is standardized with reference to fall 1976 scores on the same measure (see text for technical detail and explanation of exceptions). The means for each variable are based on cases with complete data for all timepoints shown.

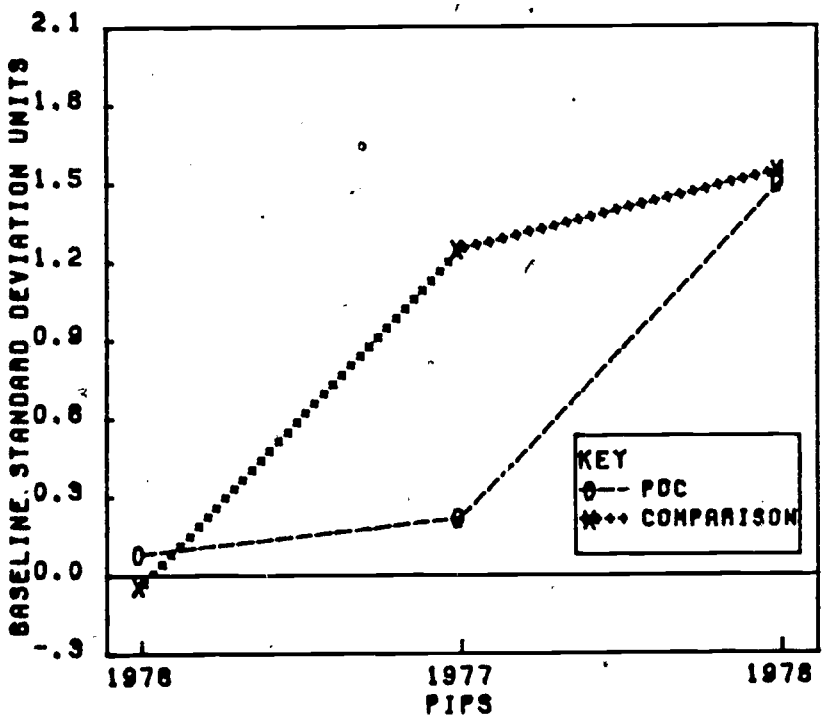
Figure 2d

Plot of PDC and Comparison Group Means on Measures in the Fall 1976, Spring 1977, and Spring 1978 Batteries (Spanish-Dominant Analytic Sample)



N(PDC) = 19

N(COMPARISON) = 26



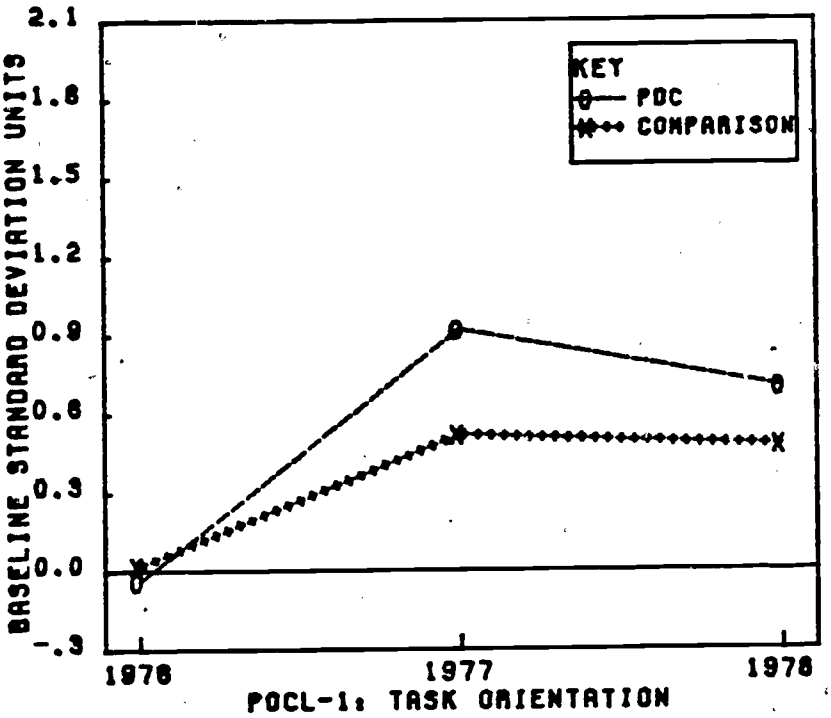
N(PDC) = 19

N(COMPARISON) = 26

Note: Each of the means plotted is standardized with reference to fall 1976 scores on the same measure (see text for technical detail and explanation of exceptions). The means for each variable are based on cases with complete data for all timepoints shown.

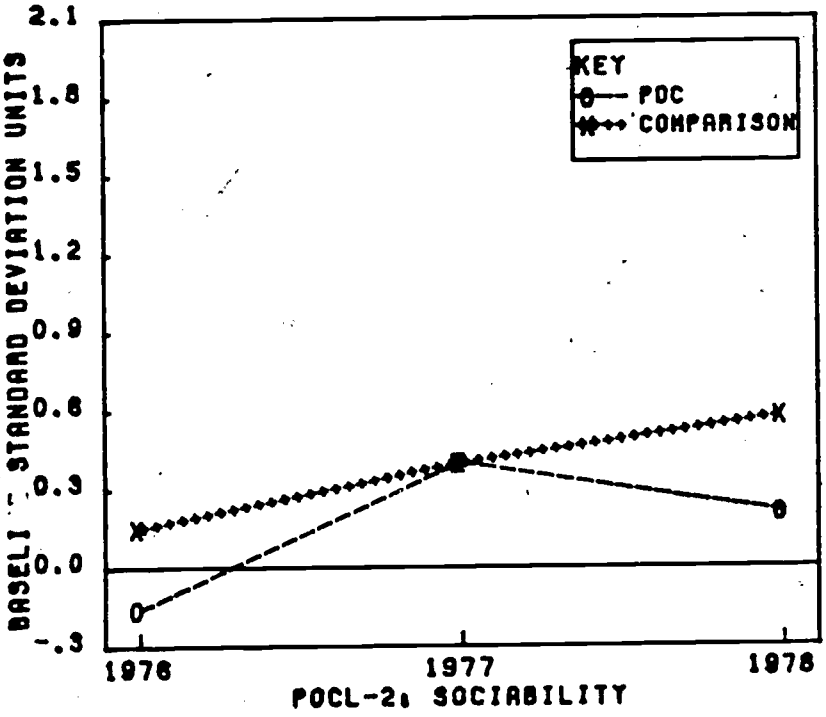
Figure 2 e

Plot of PDC and Comparison Group Means on Measures in the Fall 1976, Spring 1977, and Spring 1978 Batteries (Spanish-Dominant Analytic Sample)



N (PDC) =18

N (COMPARISON) =22



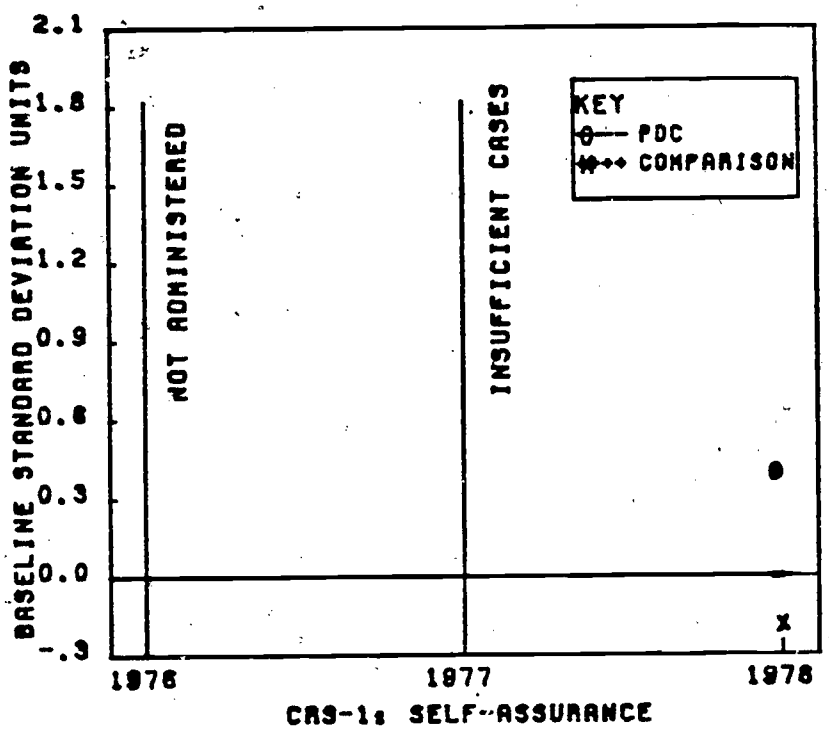
N (PDC) =18

N (COMPARISON) =22

Note: Each of the means plotted is standardized with reference to fall 1976 scores on the same measure (see text for technical detail and explanation of exceptions). The means for each variable are based on cases with complete data for all timepoints shown.

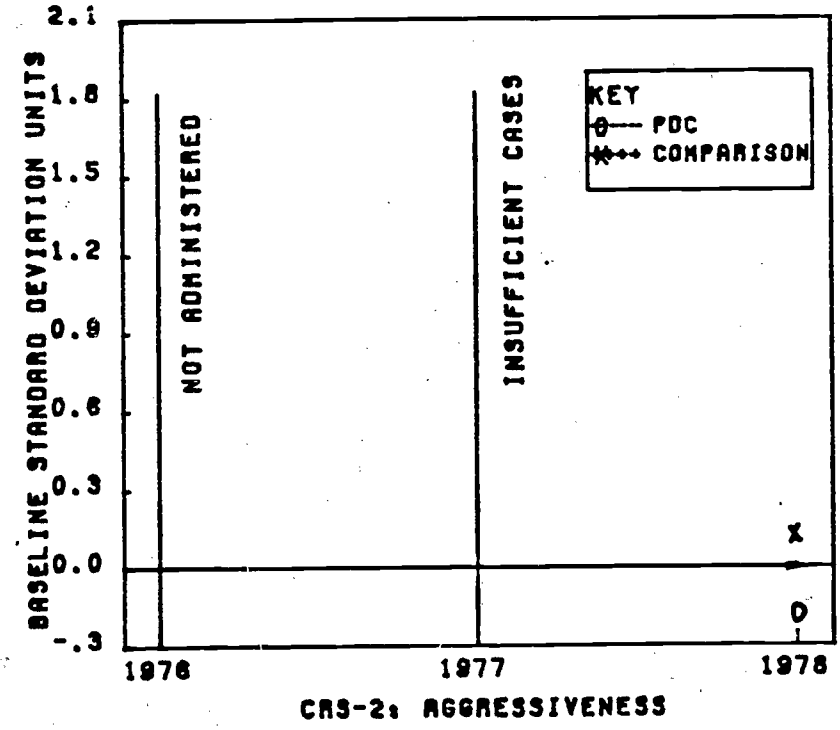
Figure 2f

Plot of PDC and Comparison Group Means on Measures in the Fall 1976, Spring 1977, and Spring 1978 Batteries (Spanish-Dominant Analytic Sample)



N (PDC) =19

N (COMPARISON) =22



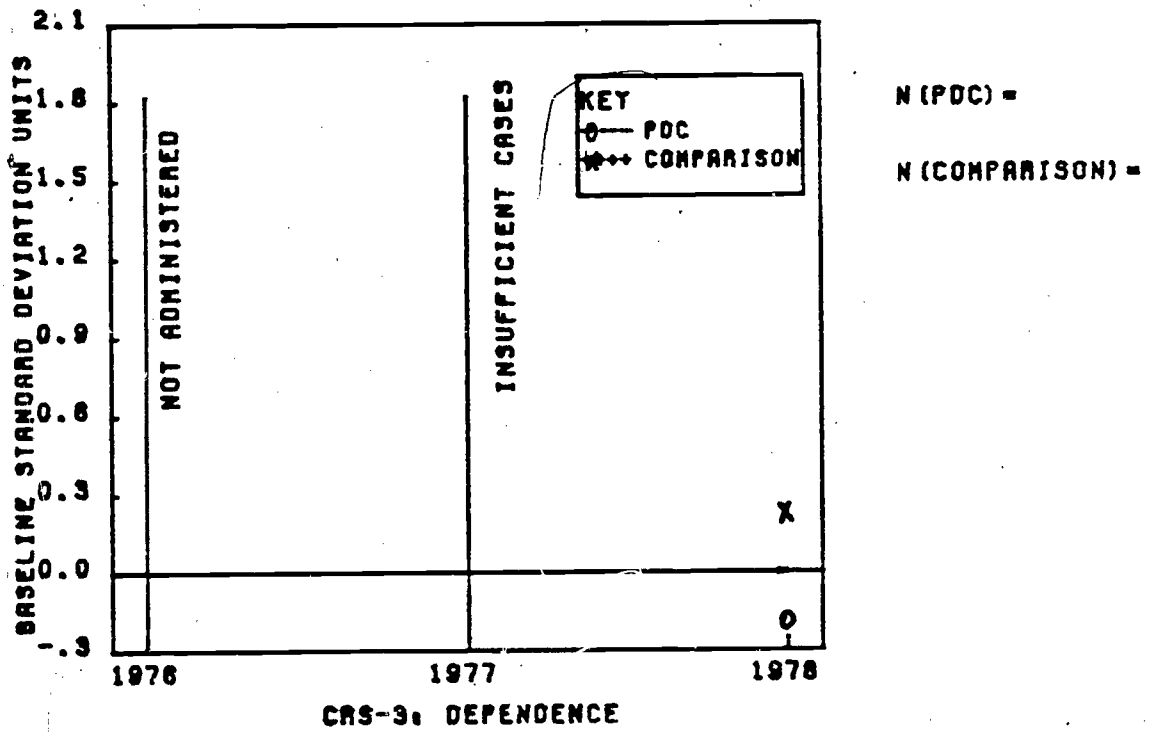
N (PDC) =20

N (COMPARISON) =26

Note: Each of the means plotted is standardized with reference to fall 1976 scores on the same measure (see text for technical detail and explanation of exceptions). The means for each variable are based on cases with complete data for all timepoints shown.

Figure 2 g

Plot of PDC and Comparison Group Means on Measures in the Fall 1976, Spring 1977, and Spring 1978 Batteries (Spanish-Dominant Analytic Sample)



Note: Each of the means plotted is standardized with reference to fall 1976 scores on the same measure (see text for technical detail and explanation of exceptions). The means for each variable are based on cases with complete data for all timepoints shown.

year-to-year growth on the measure illustrated. And since all the means are standardized, growth curves for one measure can be compared with growth curves for others (the "hill" is steeper for some). Note, however, that the group means plotted represent children's scores unadjusted for other factors (such as age, sex, or covariates related to test performance). In the analyses of program impact presented later in this report, adjustments have been made for a number of other factors extraneous to the effect of the PDC program.

Internal consistency. Table 11 presents internal consistency (Cronbach's alpha) coefficients for each measure in the spring 1978 battery. For the English-dominant analytic sample, all these coefficients continue to exceed our criterion of .65, indicating that each one remains an adequately homogeneous measure of a central construct. In the Spanish-dominant sample, coefficients have dropped below the criterion level for BSM-Spanish, Verbal Fluency, and the "Dependence" scale of the CRS. Examination of the standard deviations of these measures (refer back to Table 10) reveals that they are lower in every case for the Spanish-dominant than for the English-dominant sample. And for the Spanish-dominant children, the standard deviations of BSM-Spanish and Verbal Fluency tend to be lower than they have been in past years. Thus the low reliability coefficients may reflect a condition of diminishing variation on these measures for Spanish-dominant children. And this, in turn, may be a consequence of diminishing sample size. The small size of the Spanish-dominant sample presents a general problem that will be discussed later in this report.

Stability. Although we have established no criterion that calls for a high correlation of scores from one testing period to another, these test-retest coefficients (Table 12) add to what we know about the properties of each measure, and thus constitute useful documentation.

Relationship to "social competence." As in analyses of spring 1977 data, we examined the relationship of each measure in the spring 1978 battery to a set of measures established as proxy criteria for social competence. (Only children from the English-dominant analytic sample entered into these analyses, due to sample size considerations.) The criteria consisted of the two subscales of the POCL ("Task Orientation" and "Sociability") and the three subscales of the revised Child Rating Scale ("Self-Assurance," "Aggressiveness," and "Dependence"). The rationale for selection of these criteria was that a broad range of the child's characteristics--social, emotional, cognitive, linguistic, and psychomotor--is spanned

Table 11

Reliability of Child Measures[a]:
Cronbach's Alpha (Internal Consistency)
PDC Spring 1978 Data

Measures	English-Dominant Children		Spanish-Dominant Children	
COGNITIVE-LANGUAGE	*N*	α	*N*	α
BSM-English	*636*	.74	*36*	.85
BSM-Spanish[b]	*14*	.95	*46*	.56
Verbal Fluency	*636*	.68	*47*	.51
Verbal Memory-1	*636*	.73	*47*	.87
Verbal Memory-3	*636*	.75	*46*	.79
Draw-A-Child	*636*	.68	*47*	.75
PSYCHOMOTOR				
Arm Coordination	*636*	.72	*47*	.71
SOCIAL-EMOTIONAL				
POCL-1: "Task Orientation"	*595*	.92	*42*	.97
POCL-2: "Sociability"	*595*	.87	*42*	.93
CRS-1: "Self-Assurance"	*620*	.90	*21*	.85
CRS-2: "Aggressiveness"	*652*	.77	*22*	.84
CRS-3: "Dependence"	*649*	.78	*21*	.55

[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, and the reliability of the Classroom Observation System was determined differently.

[b]Texas and California only (bilingual/bicultural demonstration sites).

Table 12

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

Measures	English-Dominant Children				Spanish-Dominant Children			
	N	F76-S77 r_{tt}	S77-S78 r_{tt}	F76-S78 r_{tt}	*N*	F76-S77 r_{tt}	S77-S78 r_{tt}	F76-S78 r_{tt}
COGNITIVE-LANGUAGE								
BSM-English	*595*	.75	.69	.61	*6*	.64	.62	.50
BSM-Spanish	*5*	.88	.96	.94	*41*	.60	.74	.38
Verbal Fluency	*586*	.48	.52	.36	*45*	.62	.23	.11
Verbal Memory-1	*596*	.49	.49	.38	*45*	.67	.53	.42
Verbal Memory-3	*588*	.49	.33	.30	*43*	.55	.30	.41
Draw-A-Child	*587*	.56	.42	.42	*45*	.27	.27	.10
PSYCHOMOTOR								
Arm Coordination	*585*	.42	.40	.34	*45*	.31	.40	.51
SOCIAL-EMOTIONAL								
PIPS	*587*	.41	.32	.20	*45*	.11	.23	-.03
POCL-1: "Task Orientation"	*528*	.42	.36	.19	*40*	.49	.17	.26
POCL-2: "Sociability"	*528*	.44	.39	.25	*40*	.39	.14	.26
CRS-1: "Self-Assurance"[a]	*464*		.40			*insufficient sample*		
CRS-2: "Aggressiveness"[a]	*495*		.32			*insufficient sample*		
CRS-3: "Dependence"[a]	*497*		.08			*insufficient sample*		

[a]The Child Rating Scale was not administered in fall, thus a test-retest coefficient can be computed for spring time points only.

when the teacher and tester assess the child's social behavior, and that these ratings thus reflect the integration of the elements of social competence--the configuration of traits that contribute to the child's everyday effectiveness.

The procedure followed in these analyses was to regress each separate measure from the spring 1978 battery on a set of predictors consisting of the proxy social competence criteria (the scales of the POCL and CRS) plus age, sex, and ethnicity (included as covariates). Interpretation of the results focused on the degree to which the "social competence" criteria contributed to prediction of test scores beyond the prediction obtained using the covariates alone. The greater the predictive value of the proxy criteria, the greater the apparent relevance of the dependent measure to the construct of social competence.

All the measures examined bore a statistically significant relationship to the criteria, as can be seen in Table 13. Fairly strong relationships were found for Verbal Fluency, Verbal Memory-3, PIAT-Math, and PIAT-Reading. Moderate relationships were found for Verbal Memory-1 and Draw-a-Child. The criteria were rather weakly related to the remaining measures: BSM-English, Arm Coordination, and PIPS.

Correlations among measures. For the sake of further documenting the psychometric properties of the measures, Table 14 presents a matrix of intercorrelations for the English-dominant sample. Table 15 presents a corresponding matrix for the Spanish-dominant sample.

Factor structure. In Tables 16 and 17, the relationships depicted in the correlation matrices are reduced to a smaller number of common factors found to underlie the English and Spanish batteries. For the English-dominant sample, five factors emerged. The first, which might be labeled "general cognitive," consists primarily of BSM-English, Draw-a-Child, PIAT-Math, and PIAT-Reading. The second factor consists mainly of two of the CRS scales, "Self-Assurance" and "Dependence." Since the second measure is negatively loaded, the factor may essentially represent a broader measure of self-assurance than is contained in CRS-1 alone. Factor 3 is based mainly on Verbal Fluency, Verbal Memory-1, Verbal Memory-3, and PIPS. These all share an emphasis of linguistic fluency, and, indeed, "linguistic fluency" might be an appropriate name for the third factor. Factor 4 consists of the two subscales of the POCL, and perhaps represents the "halo" common to these two measures, which are aggregates of ratings assigned by the same tester. Arm Coordination and the "Aggressiveness"

Table 12

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

Test (N = 479)	% 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	.24**	.02*	Ethnicity	-.41
			POCL-2	.10
			Age	.10
Verbal Fluency	.16**	.15**	POCL-1	.19
			POCL-2	.13
			CRS-1	.12
Verbal Memory-1	.08**	.07**	POCL-1	.18
Verbal Memory-3	.15**	.11**	POCL-1	.22
			Age	.15
Arm Coordination	.17**	.02*	Sex	-.29
			Age	.18
			Ethnicity	.18
Draw-A-Child	.11**	.08**	CRS-1	.21
			Sex	.10
			Age	.09
PIPS	.04*	.04*	CRS-1	.10
PIAT-Math	.25**	.13**	Ethnicity	-.28
			POCL-1	.24
			CRS-1	.19
			Age	.16
PIAT-Reading	.21**	.15**	POCL-1	.27
			CRS-1	.20
			Ethnicity	-.20
			Age	.10

*probability of associated F ratio $\leq$ .05

**probability of associated F ratio $\leq$ .0001

KEY:

POCL-1: "Task Orientation"
POCL-2: "Sociability"
CRS-1: " Self-assurance"
CRS-2: " Aggressiveness"
CRS-3: "Dependence"

Table 14

Intercorrelations of Child Measures for English-Dominant Children[a]
PDC Spring 1978 Data

	CHILD MEASURES	BSM-English	BSM-Spanish	Verbal Fluency	Verbal Memory-1	Verbal Memory-3	Draw-A-Child	PIAT-Math	PIAT-Reading	Arm Coordination	PIPS	POCL-1	POCL-2	CRS-1	CRS-2	CRS-3
COGNITIVE-LANGUAGE	BSM-English	--	-.03 *19*	.25 *636*	.22 *636*	.18 *636*	.29 *636*	.32 *615*	.33 *600*	-.03 *636*	.08 *636*	.13 *595*	.21 *595*	.09 *553*	.00 *574*	.01 *572*
	BSM-Spanish	-.03 *19*	--	.23 *19*	.28 *19*	.23 *19*	.04 *19*	.49 *18*	.23 *16*	.20 *19*	.07 *19*	.56 *15*	.26 *15*	.11 *15*	-.18 *16*	-.06 *15*
	Verbal Fluency	.25 *636*	.23 *19*	--	.39 *636*	.39 *636*	.26 *636*	.32 *615*	.33 *600*	.14 *636*	.27 *636*	.33 *595*	.26 *595*	.25 *553*	.02 *574*	-.16 *572*
	Verbal Memory-1	.22 *636*	.28 *19*	.39 *636*	--	.34 *636*	.17 *636*	.21 *615*	.23 *600*	.01 *636*	.21 *636*	.23 *595*	.09 *595*	.19 *553*	-.07 *574*	-.10 *572*
	Verbal Memory-3	.18 *636*	.23 *19*	.39 *636*	.34 *636*	--	.20 *636*	.29 *615*	.17 *600*	.11 *636*	.18 *636*	.33 *595*	.21 *595*	.19 *553*	.01 *574*	-.14 *572*
	Draw-A-Child	.29 *636*	.04 *19*	.26 *636*	.17 *636*	.20 *636*	--	.30 *615*	.31 *600*	.04 *636*	.11 *636*	.14 *595*	.09 *595*	.28 *553*	-.02 *574*	-.15 *572*
	PIAT-Math	.32 *615*	.49 *18*	.32 *615*	.21 *615*	.29 *615*	.30 *615*	--	.53 *600*	.09 *615*	.10 *615*	.34 *575*	.24 *575*	.28 *535*	-.02 *556*	-.09 *555*
	PIAT-Reading	.33 *600*	.23 *16*	.33 *600*	.23 *600*	.17 *600*	.31 *600*	.53 *600*	--	.01 *600*	.12 *600*	.35 *561*	.20 *561*	.30 *521*	-.03 *541*	-.11 *540*
PSYCHO-MOTOR	Arm Coordination	-.03 *636*	.20 *19*	.14 *636*	.01 *636*	.11 *636*	.04 *636*	.09 *615*	.01 *600*	--	.03 *636*	.09 *595*	.01 *595*	.11 *553*	.08 *574*	-.11 *572*
SOCIAL-EMOTIONAL	PIPS	.08 *636*	.07 *19*	.27 *636*	.21 *636*	.18 *636*	.11 *636*	.10 *615*	.12 *600*	.03 *636*	--	.13 *595*	.13 *595*	.15 *553*	.03 *574*	-.11 *572*
	POCL-1: "Task Orientation"	.13 *595*	.56 *15*	.33 *595*	.23 *595*	.33 *595*	.14 *595*	.34 *575*	.35 *561*	.09 *595*	.13 *595*	--	.51 *595*	.27 *515*	-.04 *536*	-.16 *535*
	POCL-2: "Sociability"	.21 *595*	.26 *15*	.26 *595*	.09 *595*	.21 *595*	.09 *595*	.24 *575*	.20 *561*	.01 *595*	.13 *595*	.51 *595*	--	.06 *515*	.20 *536*	.00 *535*
	CRS-1: "Self-Assurance"	.09 *553*	.11 *15*	.25 *553*	.19 *553*	.19 *553*	.28 *553*	.28 *535*	.30 *521*	.11 *553*	.15 *553*	.27 *515*	.06 *515*	--	-.15 *550*	-.42 *548*
	CRS-2: "Aggressiveness"	.00 *574*	-.18 *16*	.02 *574*	-.07 *574*	.01 *574*	-.02 *574*	-.02 *556*	-.03 *541*	.08 *574*	.03 *574*	-.04 *536*	.20 *536*	-.15 *550*	--	.07 *569*
	CRS-3: "Dependence"	.01 *572*	-.06 *15*	-.16 *572*	-.10 *572*	-.14 *572*	-.15 *572*	-.09 *555*	-.11 *540*	-.11 *572*	-.11 *572*	-.16 *535*	.00 *535*	-.42 *548*	.07 *569*	--

[a]Sample size for each correlation is shown in italics.

Table 15

Intercorrelations of Child Measures for Spanish-Dominant Children[a]
PDC Spring 1978 Data

	CHILD MEASURES	BSM-Spanish	BSM-English	Verbal Fluency	Verbal Memory-1	Verbal Memory-3	Draw-A-Child	PIAT-Math	PIAT-Reading	Arm Coordination	PIPS	POCL-1	POCL-2	CRS-1	CRS-2	CRS-3
COGNITIVE-LANGUAGE	BSM-Spanish	--	-.03 *35*	.48 *46*	.43 *46*	.32 *45*	.36 *46*	.42 *26*	.37 *26*	-.11 *46*	.38 *46*	.31 *41*	.19 *41*	.23 *40*	.15 *45*	-.46 *44*
	BSM-English	- 03 *35*	--	-.03 *36*	.47 *36*	.22 *36*	-.06 *36*	.24 *26*	.30 *26*	-.17 *36*	.11 *36*	.28 *31*	.34 *31*	.02 *31*	.57 *35*	.08 *34*
	Verbal Fluency	.48 *46*	-.03 *36*	--	.28 *47*	.21 *46*	.33 *47*	.45 *26*	.39 *26*	-.16 *47*	.03 *47*	.29 *42*	.30 *42*	.05 *41*	.00 *46*	-.02 *45*
	Verbal Memory-1	.43 *46*	.47 *36*	.28 *47*	--	.42 *46*	.20 *47*	.31 *26*	.24 *26*	-.10 *47*	.22 *47*	.53 *42*	.39 *42*	.00 *41*	.43 *46*	-.10 *45*
	Verbal Memory-3	.32 *45*	.22 *36*	.21 *46*	.42 *46*	--	.01 *46*	.41 *26*	.33 *26*	.18 *46*	.44 *46*	.33 *41*	.32 *41*	-.08 *40*	.32 *45*	-.05 *44*
	Draw-A-Child	.36 *46*	-.06 *36*	.33 *47*	.20 *47*	.01 *46*	--	.22 *26*	.09 *26*	.08 *47*	.14 *47*	.12 *42*	.04 *42*	.25 *41*	.00 *46*	-.04 *45*
	PIAT-Math[b]	.42 *26*	.24 *26*	.45 *26*	.31 *26*	.41 *26*	.22 *26*	--	.44 *26*	.09 *26*	.26 *26*	.29 *31*	.32 *31*	.56 *22*	.51 *26*	.05 *25*
	PIAT-Reading[b]	.37 *26*	.30 *26*	.39 *26*	.24 *26*	.33 *26*	.09 *26*	.44 *26*	--	.17 *26*	.19 *26*	.10 *31*	.11 *31*	.36 *22*	.19 *26*	.00 *25*
PSYCHO-MOTOR	Arm Coordination	-.11 *46*	-.17 *36*	-.16 *47*	-.10 *47*	.18 *46*	.08 *47*	.09 *26*	.17 *26*	--	.06 *47*	-.40 *42*	-.46 *42*	-.08 *41*	-.05 *46*	.04 *45*
SOCIAL-EMOTIONAL	PIPS	.38 *46*	.11 *36*	.03 *47*	.22 *47*	.44 *46*	.14 *47*	.26 *26*	.19 *26*	.06 *47*	--	.36 *42*	.29 *42*	.17 *41*	.06 *46*	-.27 *45*
	POCL-1: "Task Orientation"	.31 *41*	.28 *31*	.29 *42*	.53 *42*	.33 *41*	.12 *42*	.29 *21*	.10 *31*	-.40 *42*	.36 *42*	--	.85 *42*	.30 *32*	.28 *41*	-.38 *41*
	POCL-2: "Sociability"	.19 *41*	.34 *31*	.30 *42*	.39 *42*	.32 *41*	.04 *42*	.32 *21*	.11 *31*	-.46 *42*	.29 *42*	.85 *42*	--	.29 *32*	.29 *41*	-.20 *41*
	CRS-1: "Self-Assurance"	.23 *40*	.02 *31*	.05 *41*	.00 *41*	-.08 *40*	.25 *41*	.56 *22*	.36 *22*	-.08 *41*	.17 *41*	.30 *32*	.29 *32*	--	.18 *41*	-.08 *40*
	CRS-2: "Aggressiveness"	.15 *45*	.57 *35*	.00 *46*	.43 *46*	.32 *45*	.00 *46*	.51 *26*	.19 *26*	-.05 *46*	.06 *46*	.28 *41*	.29 *41*	.18 *41*	--	-.13 *45*
	CRS-3: "Dependence"	-.16 *41*	-.08 *34*	-.02 *45*	-.10 *45*	-.05 *44*	-.04 *45*	.05 *25*	.00 *25*	.04 *45*	-.27 *45*	-.38 *41*	-.20 *41*	-.08 *40*	-.13 *45*	--

[a] Sample size for each correlation is shown in italics.

[b] Administered in English.

Table 16

Factor Analysis[a] of Scores on Child Measures, English-Dominant Children
Spring 1978

N = 479	Factor Loading of Child Measures *(highest loading italicized)*				
Child Measure	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
BSM-English	*.68*	-.20	.21	.03	-.02
Verbal Fluency	.31	.15	*.58*	.25	.04
Verbal Memory-1	.21	-.01	*.73*	.01	-.19
Verbal Memory-3	.15	.14	*.63*	.20	.06
Arm Coordination	-.09	.43	.10	.03	*.52*
Draw-A-Child	*.66*	.22	.12	-.15	.19
PIPS	-.07	.05	*.62*	.03	.11
PIAT-Math	*.65*	.15	.11	.33	-.05
PIAT-Reading	*.72*	.12	.03	.30	-.12
POCL-1: "Task Orientation"	.16	.21	.21	*.80*	-.11
POCL-2: "Sociability"	.12	-.13	.13	*.80*	.22
CRS-1: "Self-Assurance"	.26	*.73*	.11	.11	-.15
CRS-2: "Aggressiveness"	.06	-.24	-.02	.07	*.81*
CRS-3: "Dependence"	.00	-*.77*	-.11	.01	.00
% of Total Variance Accounted for	15.2	11.5	13.1	11.5	8.0

[a]Principal components solution, varimax rotation

Table 17

Factor Analysis[a] of Scores on Child Measures, Spanish-Dominant Children

Spring 1978

N = 35	Factor Loading of Child Measures *(highest loading italicized)*				
Child Measures	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
BSM-Spanish	.34	.08	*.63*	-.12	.46
Verbal Fluency	.45	.36	*.60*	-.30	-.11
Verbal Memory-1	*.69*	.03	.25	.45	.00
Verbal Memory-3	*.81*	.06	-.03	.08	.31
Arm Coordination	.28	*-.82*	.07	.13	.07
Draw-A-Child	.01	-.10	*.82*	.08	.09
PIPS	.35	-.02	.23	.06	*.67*
POCL-1: "Task Orientation"	.36	*.69*	.14	.34	.35
POCL-2: "Sociability"	.30	*.80*	.08	.30	.17
CRS-1: "Self-Assurance"	-.38	.21	*.56*	.48	.14
CRS-2: "Aggressiveness"	.20	.11	-.05	*.81*	.03
CRS-3: "Dependence"	.03	-.14	-.02	-.04	*-.87*
% of Total Variance Accounted for	17.2	16.9	15.9	11.9	14.2

[a] Principal components solution, varimax rotation.

scale of the CRS are the principal constituents of factor 5. The relationship between these two measures is not sufficiently obvious to warrant naming of the factor, but one plausible connection between the measures is an emphasis on physical maturity: the children who score highest on psychomotor tasks such as Arm Coordination are likely to be the same children who are physically ascendant, and thus "aggressive," in school.

As in the past, the factor structure found for the Spanish-dominant sample (Table 17) is similar to that for the English-dominant sample, yet not identical. However, the size of the sample available for this analysis (35 children) has reached the point at which factors are likely to be highly unstable from one replication to another. Thus interpretation of these results in isolation from other findings is inadvisable.

Characteristics of the Classroom Observation System

Background. At the outset of the PDC evaluation, High/Scope undertook to develop a time-sampling classroom observation system that focused on the behavior of individual children. The system that eventually emerged involved observing each child's classroom activities for two five-minute periods and tallying certain focal behaviors. The observation categories can be described broadly as: nature of classroom involvement, nature of verbal behavior, nature of interactions with peers, and nature of interactions with adults.

Adaptation of the Observation System as a measure of classroom atmosphere. Past analyses have revealed that the measures produced by the Observation System have only a tenuous relationship with other child measures, even when observation variables were correlated with teachers' ratings of similar dimensions. The likely explanation is that the brief sampling of classroom interactions does not adequately reflect these dimensions of a child's behavior. Thus it was decided that, in the future, observation data would be regarded as reflections of classroom atmosphere rather than as reflections of the traits of the individuals observed. In essence, what we observed among children in the classroom was to be taken as an aspect of the classroom's personality rather than the child's. (There is precedent for this approach in the literature on unit-of-analysis decisions in educational research.)

We arrived at classroom-level measures by selecting PDC and comparison classrooms in which at least three children had been observed in spring 1978 and, within each of those classrooms, averaging the values of the observation variables across children. The final sample available for analysis consisted of 39 PDC and 42 comparison classrooms.

Results of preliminary analyses. Following the pattern established for examination of other PDC measures, our first analyses of the Observation System focused on its psychometric properties. The main purpose of these analyses was to examine the relationship of observation variables to each other and to other measures in the PDC battery, also averaged across the children within each classroom. The resulting correlations were considerably lower than those that have been found for other PDC measures--so low that they cast doubt on utility of the observation variables as predictive or explanatory variables. Nevertheless, we proceeded to an analysis of the differences between PDC and comparison classrooms on these variables, more for psychometric purposes than for purposes of studying program impact. That is, since PDC and comparison classrooms are expected to differ in atmosphere, a finding of significant differences on the observation variables would suggest that the system had succeeded in portraying some important dimensions of that atmosphere. A finding of no difference on the observation variables would leave open the question of whether the null finding was due to invalidity of the measures or to a genuine absence of differences.

We performed two series of analyses of variance in examining the differences between PDC and comparison classrooms on the seven main variables produced by the observation system. In the first series, the dependent measures were analyzed in the form of raw percentages--each variable expressing the proportion of all observed occasions on which a critical event occurred (e.g., percent of the total time that children were involved in social behavior). But since percentage data tend not to conform to the homogeneity-of-variance assumption fundamental to analysis of variance, we created a second set of variables equal to:

$$(2\ (\arcsin\ (\text{original variable})^{1/2})).$$

This transformation, suggested by Winer, "is effective in stabilizing the variances"[1] of the original variable across design cells.

[1] Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1971. Pp. 399-400.

Subsequent analyses of both the transformed and untransformed sets of variables revealed no differences whatever between PDC and comparison classrooms. Thus, since it cannot be determined whether this null finding is due to the actual absence of classroom differences or simply to the inability of the observation system to detect such differences, the analyses will not be further documented.

From spring 1979 onward, a new observation system will be used in the PDC evaluation. This system, to be described in future reports, is expected to provide greater insight into classroom processes and greater measurement reliability than the system employed through spring of 1978.

IV

EXAMINATION OF TREATMENT EFFECTS

Guiding Questions

There are many facets to the question, "Has PDC had an impact on children?" Each facet gives voice to a different emphasis:

- *Has PDC had the same impact on children at all sites?*
- *Is PDC's impact evidenced equally on all measures?*
- *Is PDC's impact the same for all children, regardless of background?*
- *Has PDC had an impact since the beginning of the Head Start year? Since the beginning of the kindergarten year?*

Clearly there can be no single answer to the question of PDC's impact, since it is not really a single question. Thus we took several different approaches to analysis of spring 1978 data, each approach covering a somewhat different aspect of the central issue.

Analytic Design

The basic design of the impact analyses reported here was similar to the design used for analysis of impact through the Head Start year: by means of analysis of covariance, spring 1978 test scores were each adjusted for prior scores on the same measures (except in the repeated measures analyses), for fall 1976 WPPSI[1] score, and for the factors of site, ethnicity, sex, and prior preschool experience. Each analysis

[1]The Block Design subtest of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) was administered to all children at the time of program entry (fall 1976) as a baseline measure of general cognitive aptitude. This fall 1976 measure continues to serve in our analyses as a covariate representing one dimension of children's initial ability.

then yielded an assessment of the difference between PDC and comparison groups on the adjusted score. This design allowed evaluation of both the main effect of the PDC program and the interaction of the program effect with site conditions. That is, we examined first the question of whether PDC had, on the whole, produced an effect for PDC children relative to comparison children, then the question of whether this effect was the same at all sites.

In concise terms, the analytic model represents a proposition that...

An individual child's status on a given measure in spring 1978 = The sum of the contributions of:

Grand mean
Mean spring 1978 status of all children on the measure under analysis.

\+ Covariates
Baseline performance status (fall 1976 WPPSI score and, in some analyses, fall 1976/spring 1977 status on the measure under analysis).

\+ Program factor
The child's educational program (PDC or comparison).

\+ Site factor
Conditions associated with individual sites.

\+ Sex factor
Conditions associated with being a boy or a girl.

\+ Preschool factor
Preschool experience prior to the 1976-1977 Head Start year.

\+ Program-by-site interaction term
Conditions associated with local programs as they differed from national programs.

\+ Error term
Variation within the categories defined by this design, due to factors not specified, plus variation due to unspecified interactions of the design factors.

For each of the two analytic samples of children (English-dominant and Spanish-dominant), three separate analyses were carried out for each measure in the spring 1978 battery[1]--a total of 68 analyses. The three series of analyses all fit the framework just described but addressed three somewhat different questions:

1. *Has a group difference arisen since fall of the Head Start year?*

   (Approach: analysis of covariance with two covariates and one dependent variable.)[2]

2. *Has a group difference arisen since spring of the Head Start year?*

   (Approach: analysis of covariance with three covariates and one dependent variable.)[2]

3. *Is there a difference in the growth curves of the two groups?*

   (Approach: repeated measures analysis of covariance with one covariate and three simultaneous dependent variables.)[2]

The first question is probably the primary one for the PDC evaluation. For the respective kindergarten-year measures analyzed, each child's fall 1976 score on the same measure served as a covariate, along with the child's fall 1976 WPPSI score. The analysis then focused upon change in the dependent measure since fall 1976, allowing estimation of PDC children's gains relative to comparison children during the Head Start and kindergarten years.

The second question is logically subsidiary to the first. It basically asks whether PDC-versus-comparison differences have become manifest during the kindergarten year. If they have,

---

[1]Due to the small size of the Spanish-dominant analytic sample (N=46) it was necessary to reduce the number of parameters involved in the analyses. Thus the design for this sample excluded the factors of site, ethnicity, sex, and prior preschool experience.

[2]Because the initial match between PDC and comparison groups was quite close, as noted in this and past reports, the purpose of using covariates to "adjust" scores is not so much to equalize the initial status of the PDC and comparison groups as to lend greater precision to the analyses, by reducing the amount of variation due to unspecified "error."

then, theoretically, the differences should be reflected in analyses addressing the first question as well, since those analyses tell whether differences have arisen at any time over the course of the program. However, since the second series of analyses introduced a new covariate (spring 1977 score on the dependent measure) not present in the first series, the statistical result could conceivably indicate a group difference for the kindergarten year when none was indicated for the Head Start and kindergarten years together.

The third question, concerning growth curves, was addressed by performing repeated measures analyses of covariance. The results of these analyses should be positive if there are differences between PDC and comparison children (after adjustment for extraneous factors) at any point from fall 1976 to spring 1978. Logically, one would expect these analyses to indicate a PDC-comparison difference for any measure that reflected a group difference in the first or second series of analyses. However, because we interpreted these analyses conservatively[1], it is possible for the results to appear somewhat inconsistent with the results of the preceding analyses.

---

[1]One of the assumptions involved in univariate repeated measures analysis is that all possible pairs of the repeated measures are equally correlated. (This is known as the assumption of "compound symmetry" of correlations.) In PDC's case, as in the case of most extended longitudinal studies, this assumption often is not sustained by the data. Thus, in examining the results of the repeated measures analyses, we chose not to interpret the evidence bearing on the equivalence or non-equivalence of the slopes of PDC and comparison group growth curves. Instead, we based our conclusions about program effects on that part of each repeated measures analysis that tested the difference between the PDC and comparison means for all repetitions of the measure analyzed (e.g.: (fall 1976+spring 1977+spring 1978 score)/3)). This meant that, in the interest of statistical prudence, we examined only one of the two ways in which the growth curves of the groups might differ. In the future, we will have access to computer programs that will circumvent the compound symmetry assumption by performing repeated measures analyses in a multivariate mode. This will allow us to deal with the question that we had to bypass here: the question of a possible difference in the slopes of the PDC and comparison curves.

Results of the Impact Analyses[1,2]

The English-dominant analytic sample. The results of the three series of analyses for the English-dominant sample are summarized in Table 18. The prevailing pattern evident there is one of no significant differences between PDC and comparison groups: the first series of analyses revealed no overall differences ("main effects," in ANOVA terms) on any measure, and the second and third series each indicated a marginally significant overall difference on just a single measure. There are more numerous indications of group differences local to particular sites ("program-by-site interactions"), but the three series of analyses are not in close agreement on these. On only one measure out of the 12 commonly examined do the three analyses consistently reflect a significant difference. That measure is the "Aggressiveness" subscale of the Child Rating Scale. Analyses 1 and 2 indicate that there are PDC-comparison differences in "Aggressiveness" at some but not all sites, while analysis 3 indicates a significant overall group difference and no site-level departures from that effect.

Figures 3a through 3g portray the relative status of PDC and comparison groups on every measure receiving a "Yes" anywhere in Table 18. Each of these figures shows, first, the status of all PDC and all comparison children combined. Then, where there was any indication of a group-by-site interaction, each figure shows plots of group means for the four sites where groups were found to depart most substantially from the overall pattern. Some insight into the nature of an interaction can be obtained by comparing the plot for each site with the plot for all sites combined. For example, in Figure 3a it can be seen that, overall, PDC children's scores on Verbal Memory-1 are lower than those of comparison children at all three time-points. In Utah and West Virginia, however, PDC children tend

---

[1]Because of limitations inherent in the computer program used for these analyses (program P2V of UCLA's BMD series), it was necessary to exclude Georgia children from all the analyses reported here. This was due to the program's inability to tolerate the absence of a comparison group in Georgia. Our imminent acquisition of a more tolerant, more general, program will eliminate this problem in the future, allowing inclusion of Georgia data.

[2]Appendix B documents the significance levels associated with the contributions of each covariate and design factor in each of the major analyses described here.

Table 18

Summary of Analyses of Program Effects for the English-Dominant Analytic Sample

Analytic Question:		Analysis #1: *"Has a group difference arisen since fall of the Head Start year?"*		Analysis #2: *"Has a group difference arisen since spring of the Head Start year?"*		Analysis #3: *"Is there a difference in the growth curves of the two groups?"*	
Design Factors:		Treatment group, site, ethnicity, sex, prior preschool experience		Treatment group, site, ethnicity, sex, prior preschool experience		Treatment group, site, ethnicity, sex, prior preschool experience	
Covariate(s):		F76 WPPSI, F76 status on the target measure[c]		F76 WPPSI, F76 status on the target measure[c], S77 status on the target measure		F76 WPPSI	
Dependent Variable(s):		S78 status on the target measure		S78 status on the target measure		F76, S77, and S78 status on the target measure	
Target Measure	Sample Size[d]	All Sites Combined	Individual Sites	All Sites Combined	Individual Sites	All Sites Combined	Individual Sites
BSM-English	568	No	No	No	No	No	No
Verbal Fluency	559	No	No	No	No	No	No
Verbal Memory-1	569	No	No	No	No	No	Yes*
Verbal Memory-3	561	No	Yes**	No	Yes**	No	No
Arm Coordination	558	No	No	No	No	No	No
Draw-a-Child	560	No	No	No	No	No	No
PIPS	560	No	No	No	No	No	No
POCL-1:"Task Orientation"	504	No	Yes*	No	No	No	No
POCL-2:"Sociability"	504	No	No	No	No	No	No
PIAT-Math[a]	586	No	Yes**	------Analyses not applicable------			
PIAT-Reading[a]	572	No	Yes**	------Analyses not applicable------			
CRS-1:"Self-Assurance"[b]	420	No	No	Yes*	No	No	No
CRS-2:"Aggressiveness"[b]	449	No	Yes**	No	Yes*	Yes*	No
CRS-3:"Dependence"[b]	451	No	No	No	No	No	No

[a]Administered in S78 only. [b]Administered in S77 and S78 only. [c]If administered in F76.

[d]Number of PDC and comparison children with complete data on all variables involved.

*p < .05 (two-tailed)
**p < .01 (two-tailed)

Figure 3a

Illustration of Main Program Effects and Program-by-Site Interactions for Verbal Memory - 1
(English-Dominant Analytic Sample)

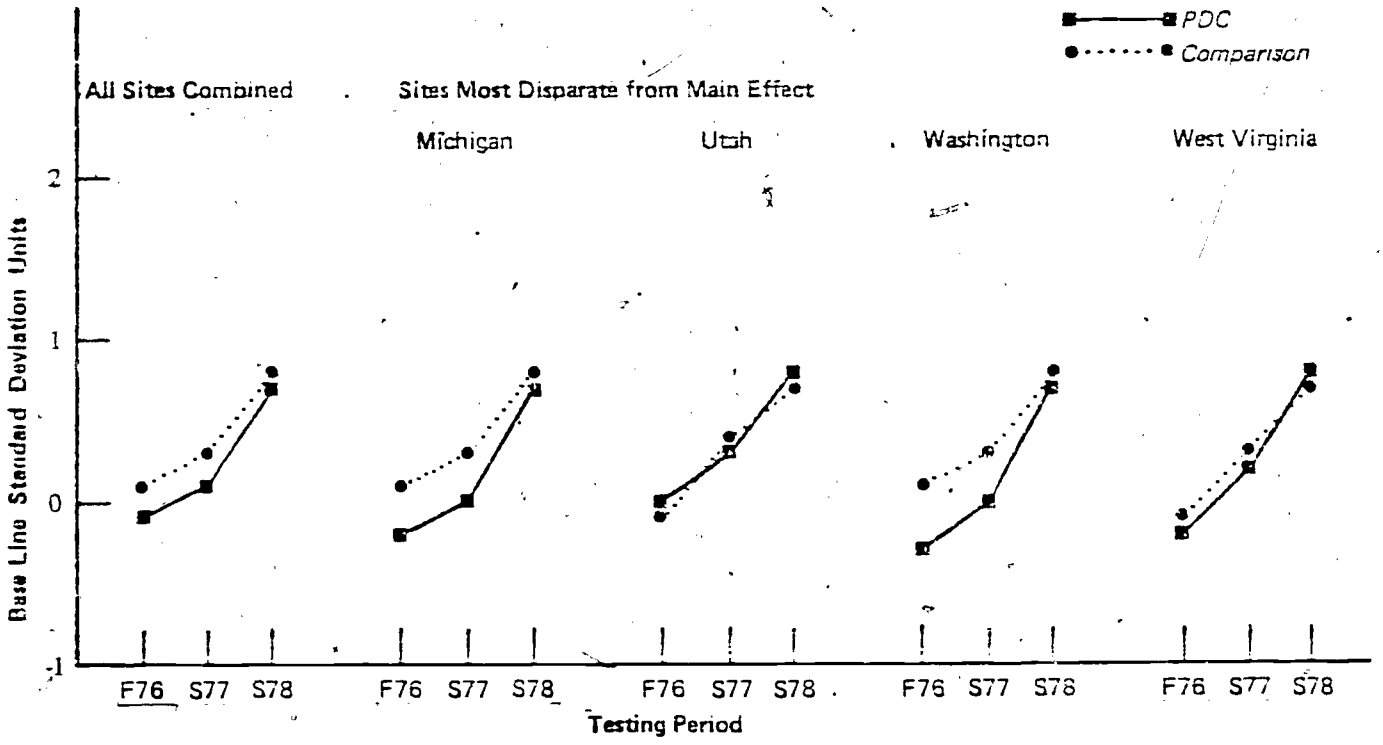
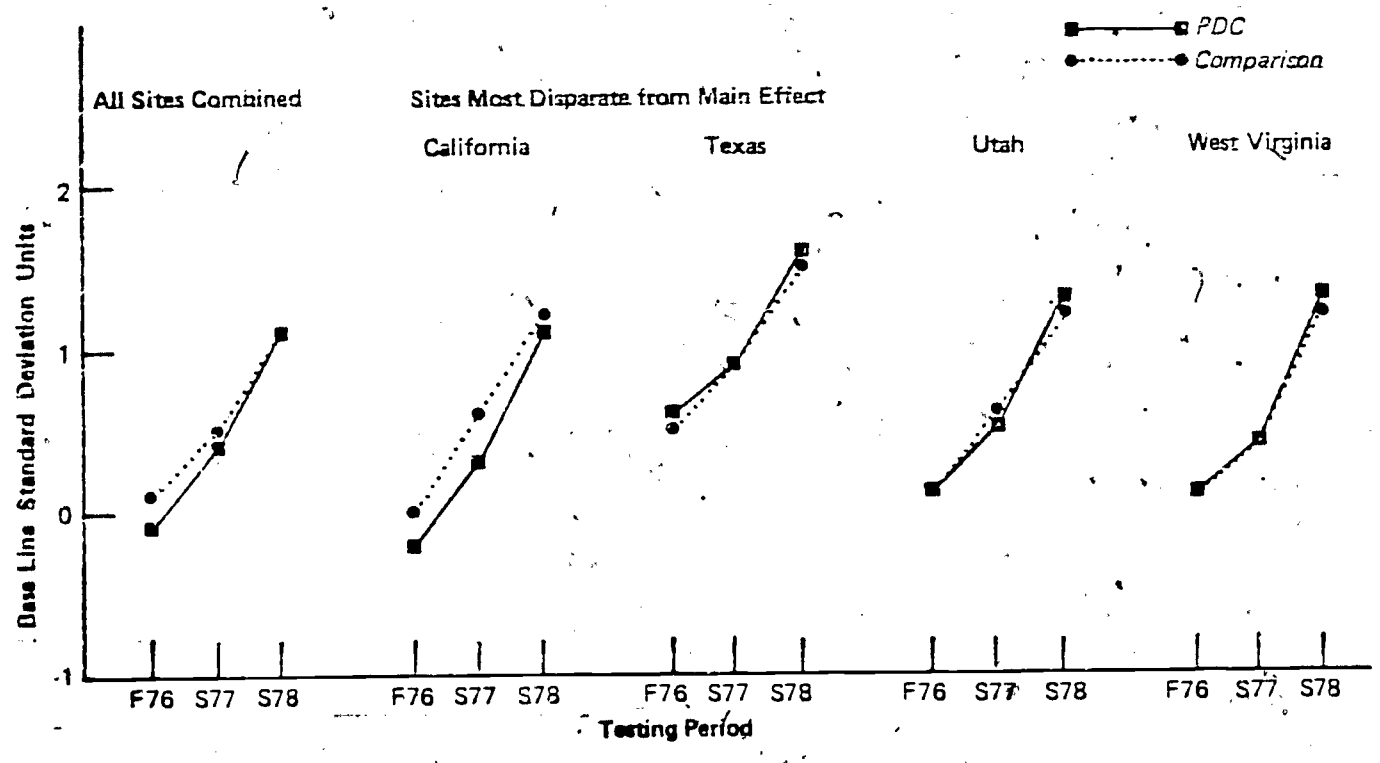


Figure 3b

Illustration of Main Program Effects and Program-by-Site Interactions for Verbal Memory - 3
(English-Dominant Analytic Sample)



Notes: The points graphed represent group means adjusted for fall 1976 WPPSI score and for the effects of sex, ethnicity, and prior preschool experience. The means are standardized on the basis of the overall mean and standard deviation for the earliest testing period shown.

Figure 3c

Illustration of Main Program Effects and Program-by-Site Interactions for POCL-1: "Task Orientation" (English-Dominant Analytic Sample)

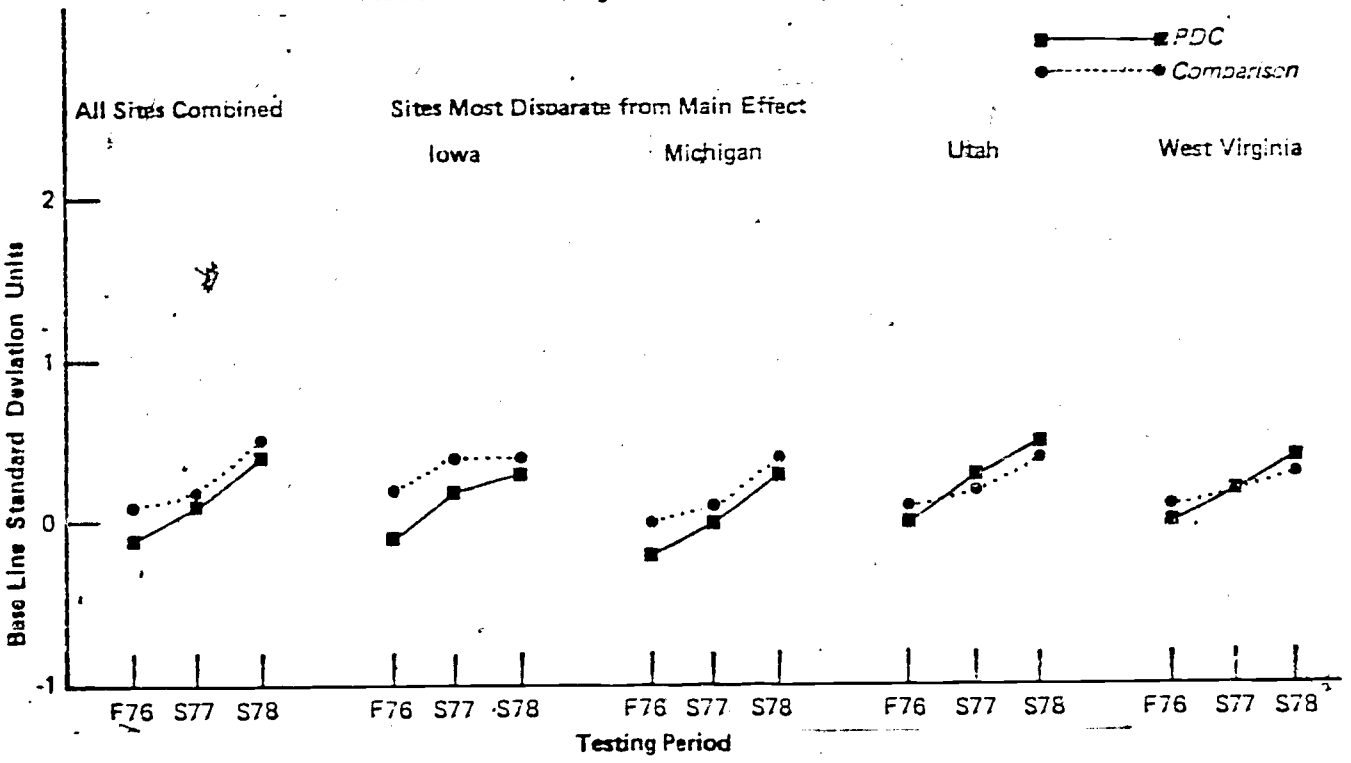
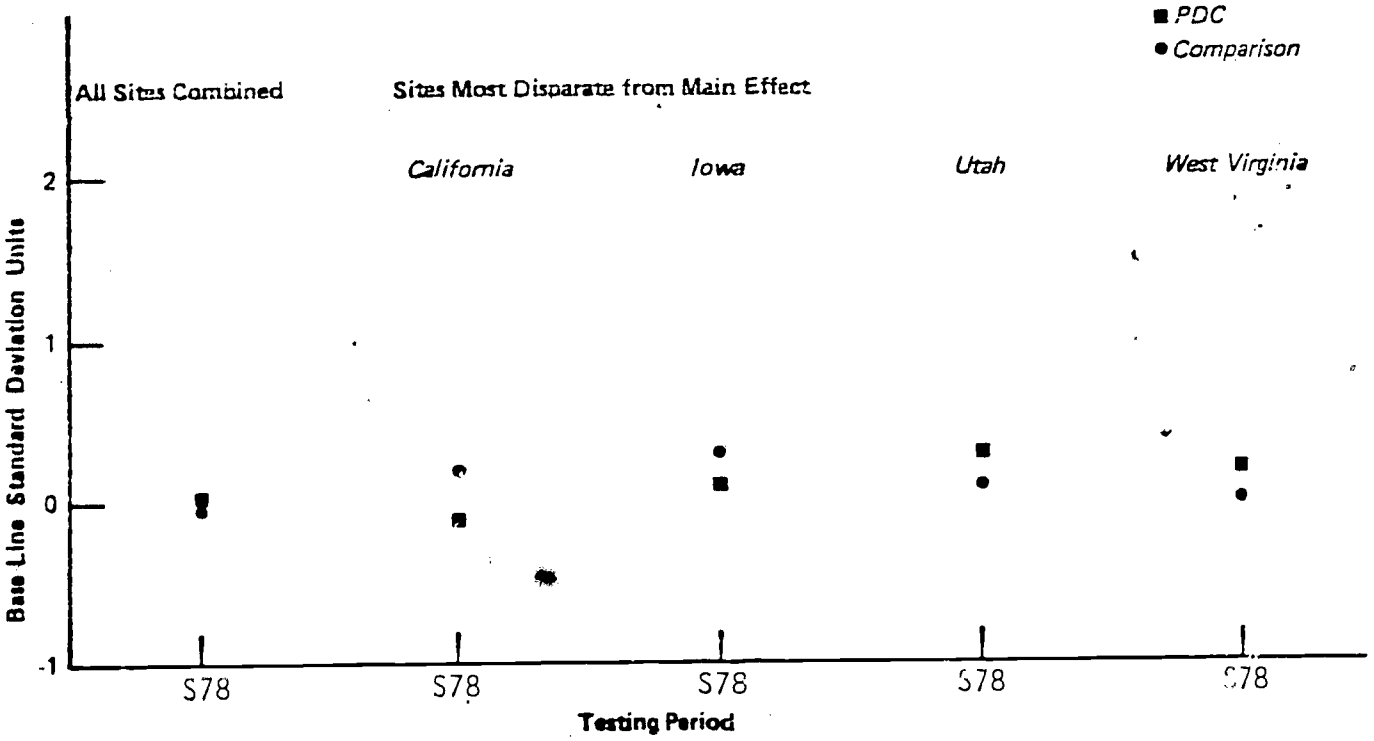


Figure 3d

Illustration of Main Program Effects and Program-by-Site Interactions for PIAT-Math (English-Dominant Analytic Sample)



Note: The points graphed represent group means adjusted for fall 1976 WRPSI score and for the effects of x, ethnicity, and prior preschool experience. The means are standardized on the basis of the overall mean and standard deviation for the earliest testing period shown.

Figure 3e

Illustration of Main Program Effects and Program-by-Site Interactions for PIAT-Reading
(English-Dominant Analytic Sample)

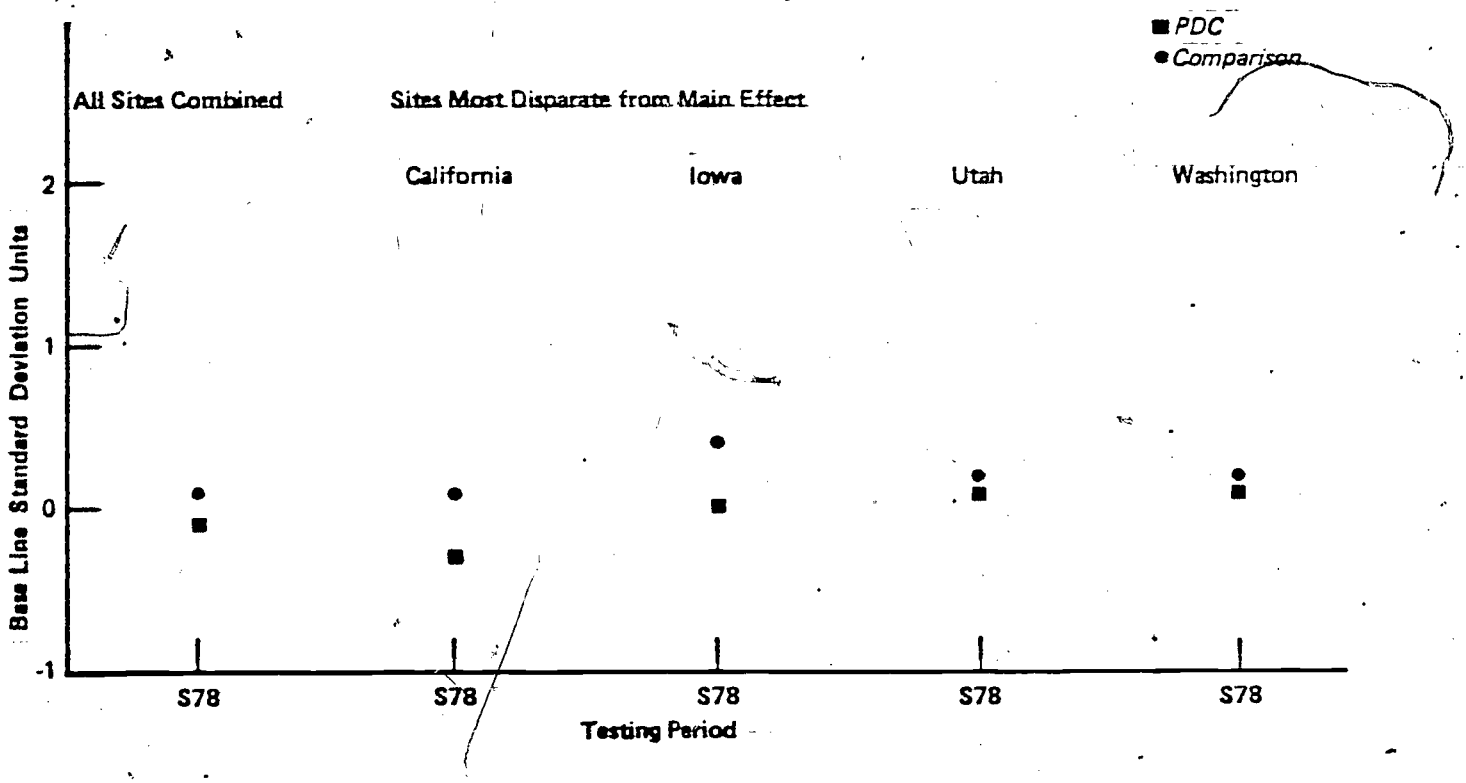
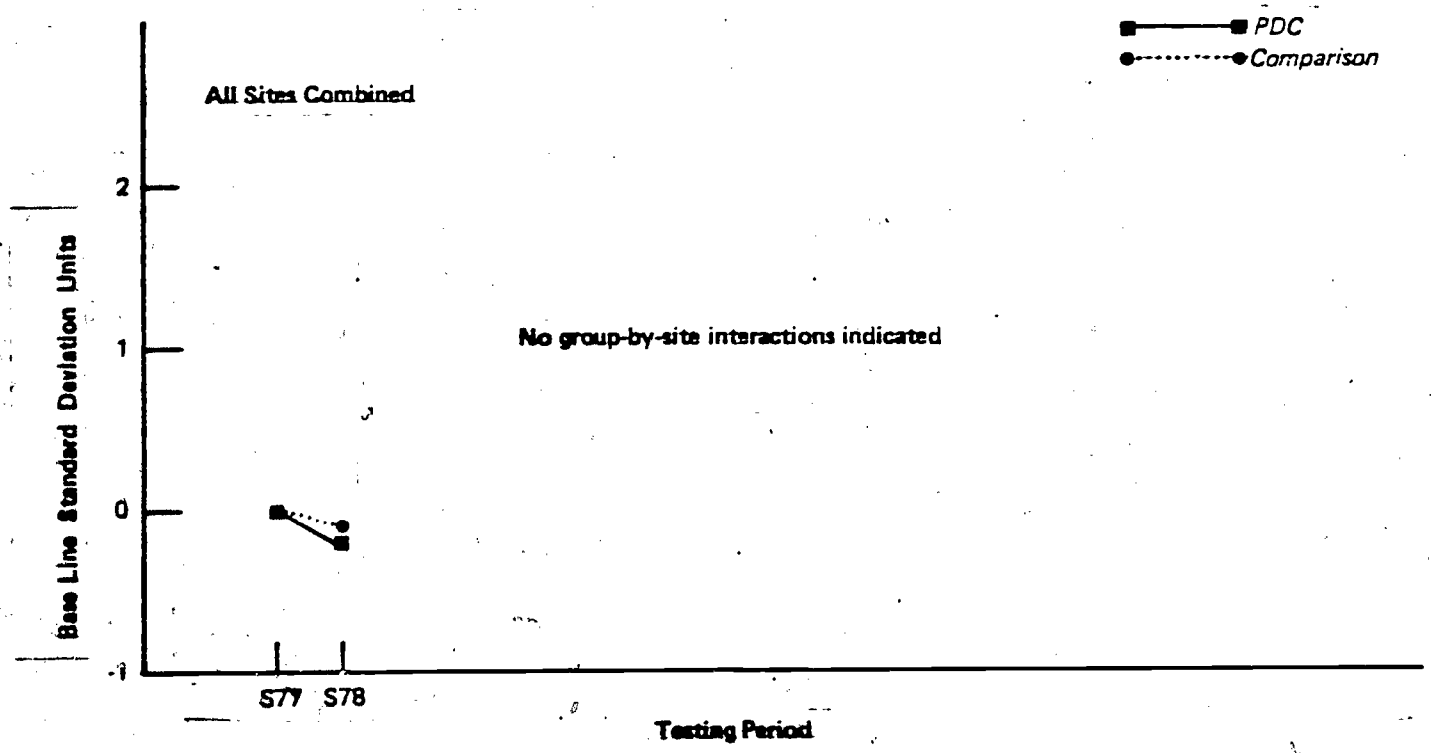


Figure 3f

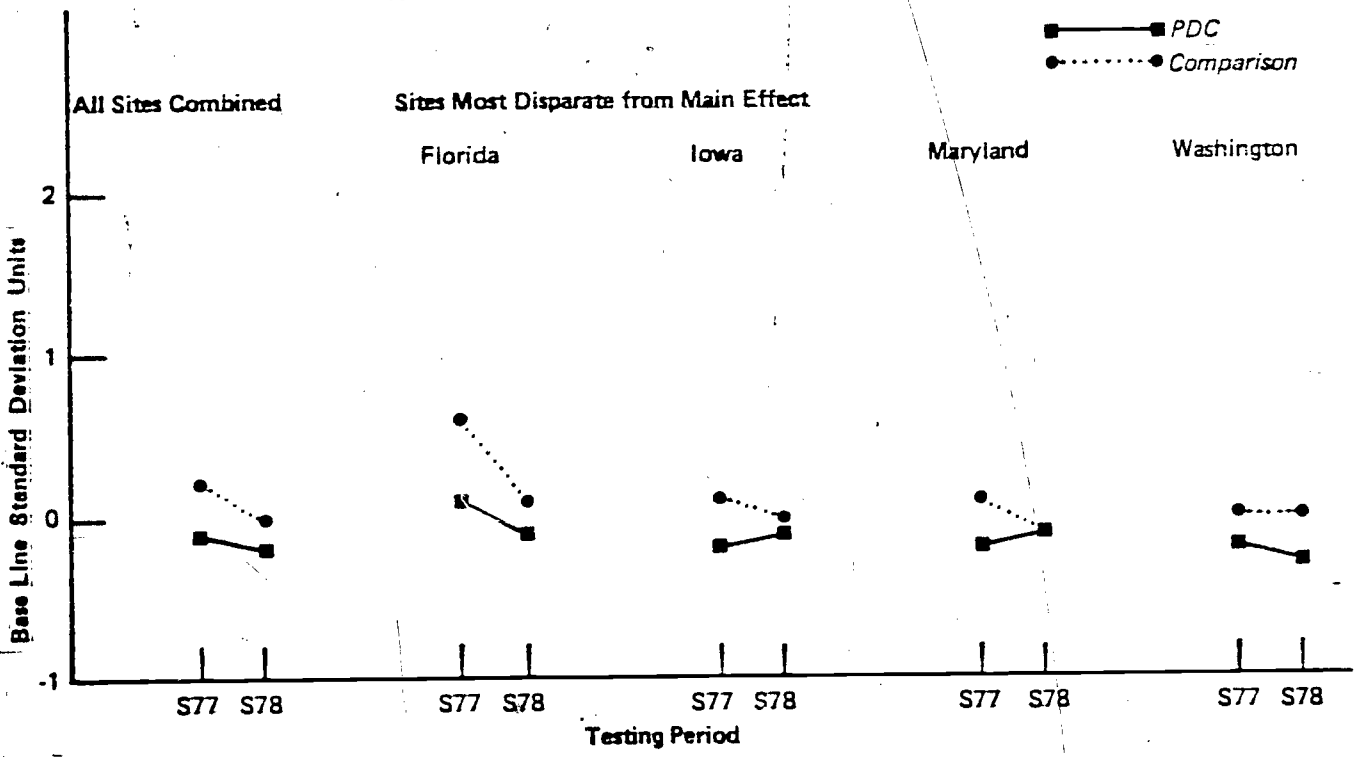
Illustration of Main Program Effects for CRS-1: "Self-Assurance"
(English-Dominant Analytic Sample)



Note: The points graphed represent group means adjusted for fall 1976 WPPSI score and for the effects of sex, ethnicity, and prior preschool experience. The means are standardized on the basis of the overall mean and standard deviation for the earliest testing period shown.

Figure 3g

Illustration of Main Program Effects and Program-by-Site Interactions for CRS-2: "Aggressiveness" (English-Dominant Analytic Sample)



Note: The points graphed represent group means adjusted for fall 1976 WPPSI score and for the effects of sex, ethnicity, and prior preschool experience. The means are standardized on the basis of the overall mean and standard deviation for the earliest testing period shown.

to be doing relatively better, while in Michigan and Washington they're doing relatively worse. It's interesting to note that Utah and West Virginia show up positively in three other figures: 3b (Verbal Memory-3), 3c (POCL-1), and 3d (PIAT-Math).

The edge that the PDC groups in Utah and West Virginia appear to have over PDC groups in other sites stands out against a background of findings that otherwise tend to be either null or inconsistent. As far as the English-dominant sample goes, only a few other findings are worthy of note. First, in no site is the PDC group's adjusted score on PIAT-Reading higher than the comparison group's. Although the overall group difference on this measure is not statistically significant, the consistent PDC decrement bears noting just because of the weight inevitably attached to reading achievement. And, although we performed no post-hoc significance tests of group differences at the site level[1], Figure 3e suggests that PDC children in California and Iowa score substantially lower in reading than their comparison group counterparts, even after adjustments have been made for WPPSI score and other design factors. PDC children also score lower, overall, on subscale 2 of the Child Rating Scale: "Aggressiveness." Figure 3g depicts PDC and comparison status on this measure for all sites in aggregation and for the four sites that differ most from the main effect. For all but one of the illustrated data points, comparison children score higher in "Aggressiveness" than PDC children (in the case of the one exception, the groups are about equal). Although it is on this measure that PDC and comparison children differ most convincingly, it is not clear what meaning can be attached to the difference. Is aggressiveness a good or a bad quality for kindergarten children? There are few clues to be found in Table 14, which shows that this scale correlates very weakly with other measures in the battery. And even if it is possible to make a pronouncement on the value of aggressiveness, is it a quality of behavior that PDC can be expected to influence? It is possible that group differences on the measure reflect something other than the effect of the PDC program. Considering the low incidence of significant group differences reflected by the 68 analyses performed for the English-dominant sample, the indications of a group difference on "Aggressiveness" may simply be due to chance.

---

[1]In the future, we intend to evaluate apparent site-level differences statistically and not just graphically. For the present, any conclusions about group differences at a given site are only conjectural.

In sum, the analyses based upon data for the English-dominant sample provide no evidence of a general PDC-favoring effect at this point. However, some of the analyses hint that PDC children at some sites may be making gains relative to comparison children while PDC children at some other sites may be losing ground. Table 19 provides some insight into the developing patterns. That table shows PDC-versus-comparison differences on the same measures graphed in Figures 3a to 3g (but now adjusted for fall 1976 status) as being either positive or negative (i.e., PDC mean higher than comparison mean or vice-versa). The magnitude of these differences has not been evaluated statistically, and it is likely that most would not reach statistical significance if they were evaluated. Nonetheless, since chance should favor the PDC and comparison groups equally, one would expect the plus and minus signs to cancel out if there were truly no differences between groups. However, at the aggregate level (all sites combined), PDC means exceed comparison means in only 25% of the instances shown.[1] Note, though, that at some sites the percentage is considerably higher while at others it is considerably lower. The sites seem to fall into four ranks with respect to the relative statuses of their PDC and comparison groups. Proceeding downward in order of PDC "advantage," the ranks are:

- Utah (58% PDC-favoring differences)
- West Virginia (50%)

- Connecticut (42%)
- Maryland (42%)
- Texas (42%)

- California (25%)
- Colorado (25%)
- Washington (25%)

- Florida (8%)
- Iowa (8%)
- Michigan (0%)

[1] In the context of this table, a high "Aggressiveness" rating is considered positive, just as is a high PIAT-Reading score. This assumption is arguable, but even if the opposite assumption is made, it has little effect on the relative status of the sites. It does, however, affect the overall proportion of PDC-favoring differences: if a low score on that scale were considered positive, the percentage of PDC-favoring differences would rise from 25 to 42.

Table 19

Relative Status of English-Dominant PDC and Comparison Groups on Selected Measures
(Adjusted for Pre-Existing Differences[a])

	Verbal Memory-1		Verbal Memory-3		POCL-1: "Task Orientation"		PIAT-Math	PIAT-Reading	CRS-1: "Self-Assurance"		CRS-2: "Aggressiveness"		% of Positive Differences (PDC > Comparison)
	S77	*S78*	*S77*	*S78*	*S77*	*S78*	*S78*	*S78*	*S77*	*S78*	*S77*	*S78*	
All Sites Combined	-	-	-	+	-	-	-	-	+	+	-	-	25
California	-	-	-	-	+	+	-	-	b	b	b	b	25
Colorado	-	-	-	+	+	-	-	-	+	-	-	-	25
Connecticut	+	+	-	-	+	-	+	-	+	-	-	-	42
Florida	-	-	-	+	-	-	-	-	-	-	-	-	8
Iowa	-	-	-	+	-	-	-	-	-	-	-	-	8
Maryland	+	+	+	+	-	-	-	-	+	-	-	-	42
Michigan	-	-	-	-	-	-	-	-	-	-	-	-	0
Texas	-	+	-	+	-	-	+	-	b	b	b	b	38
Utah	+	+	-	+	+	+	+	-	+	-	-	-	58
Washington	-	-	-	-	-	-	+	-	+	-	-	-	25
West Virginia	+	+	+	+	-	-	+	-	+	-	-	-	50
% of Positive Site-Level Differences (PDC > Comparison)	36	45	18	73	36	18	45	0	78	11	0	0	

[a]Group means adjusted for F76 WPPSI, F76 status on the dependent variable (where applicable), ethnicity, sex, and prior preschool experience.

[b]Data insufficient for analysis.

Key
+ PDC Mean > Comparison Mean
- Comparison Mean > PDC Mean

The Spanish-dominant analytic sample. The results of impact analyses for the Spanish-dominant analytic sample are presented in Table 20. That table, interpreted alongside Figures 4a through 4e, indicates that PDC children in this sample scored higher than comparison children on Arm Coordination and on POCL-1: "Task Orientation." A number of group-by-site interactions are also indicated, but these are suspect due to the small size of the PDC sample in California (only four children). Indeed, the size of the total Spanish-dominant sample is so small that even indications of main effects in these analyses do not warrant intensive interpretation. If ACYF concurs with this judgment, it would then be advisable to reconsider the conception of the evaluation for the Spanish-dominant analytic sample. Even now, as noted, degrees-of-freedom restrictions require a reduction in the complexity of the design framework used in analysis of data for the Spanish-dominant children. Given the certainty of continued attrition, it is likely that in the future further departures from the main analytic model will be necessary.

Table 20

Summary of Analyses of Program Effects for the Spanish-Dominant Analytic Sample

Analytic Question:		Analysis #1: *"Has a group difference arisen since fall of the Head Start year?"*		Analysis #2: *"Has a group difference arisen since spring of the Head Start year?"*		Analysis #3: *"Is there a difference in the growth curves of the two groups?"*	
Design Factors:		Treatment group, site		Treatment group, site		Treatment group, site	
Covariate(s):		F76 WPPSI, F76 status on the target measure[c]		F76 WPPSI, F76[c] and S77 status on the target measure		F76 WPPSI	
Dependent Variable(s):		S78 status on the target measure		S78 status on the target measure		F76, S77, and S78 status on the target measure	
Target Measure	Sample Size[b]	All Sites Combined	Individual Sites	All Sites Combined	Individual Sites	All Sites Combined	Individual Sites
BSM-Spanish	41	No	No	No	No	No	No
Verbal Fluency	45	No	No	No	No	No	No
Verbal Memory-1	45	No	No	No	No	No	No
Verbal Memory-3	43	No	No	No	No	No	No
Arm Coordination	45	Yes*	Yes*	Yes**	No	Yes*	No
Draw-a-Child	45	No	No	No	No	No	No
PIPS	45	No	Yes*	No	Yes*	No	Yes*
POCL-1:"Task Orientation"	40	Yes*	Yes**	Yes*	Yes**	No	No
POCL-2:"Sociability"	40	No	Yes**	No	Yes**	No	No
CRS-1:"Self-Assurance"[a]	41	No	No	------Analyses not applicable------			
CRS-2:"Aggressiveness"[a]	46	No	No	------Analyses not applicable------			
CRS-3:"Dependence"[a]	45	No	Yes*	-----Analyses not applicable------			

[a]Measures available for S78 only.

*p < .05 (two-tailed)
**p < .01 (two-tailed)

[b]Number of PDC and comparison children with complete data on all variables involved.

[c]If administered in F76.

Figure 4a

Illustration of Main Program Effects and Program-by-Site Interactions for Arm Coordination (Spanish-Dominant Analytic Sample)

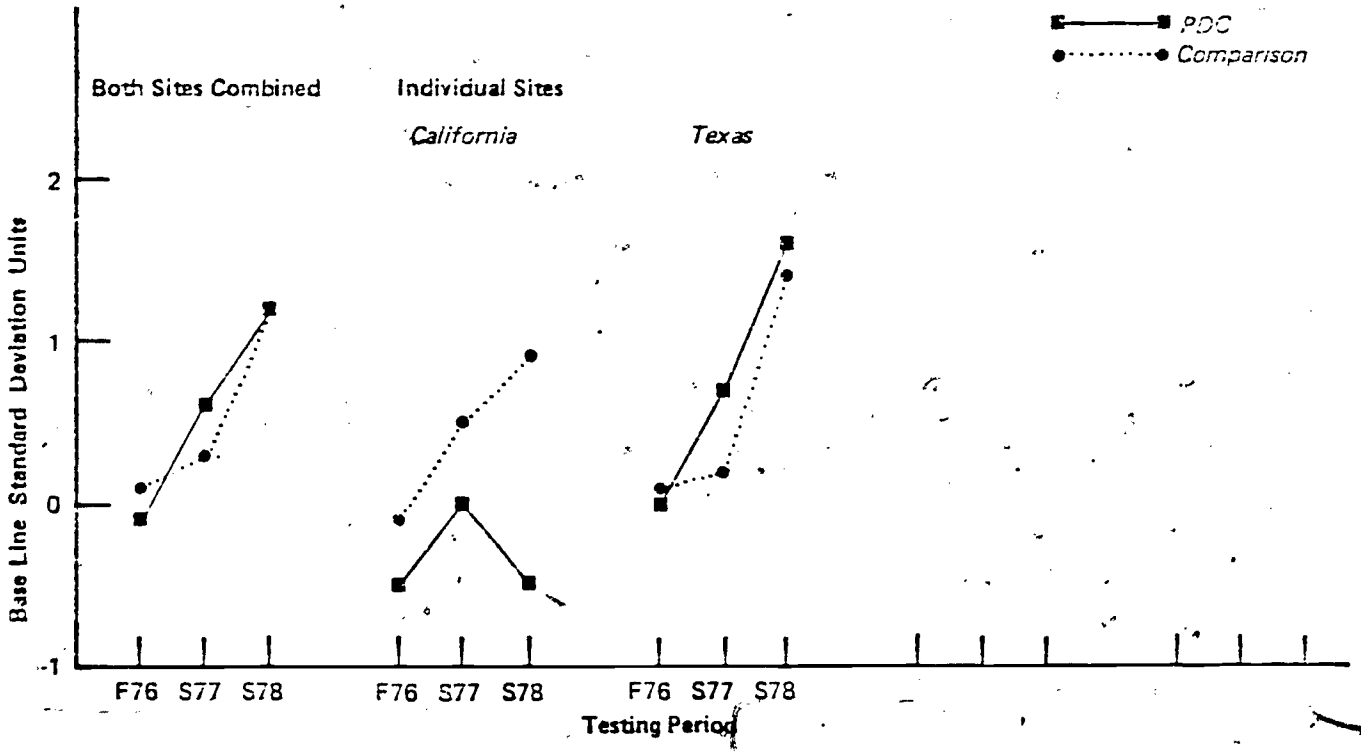
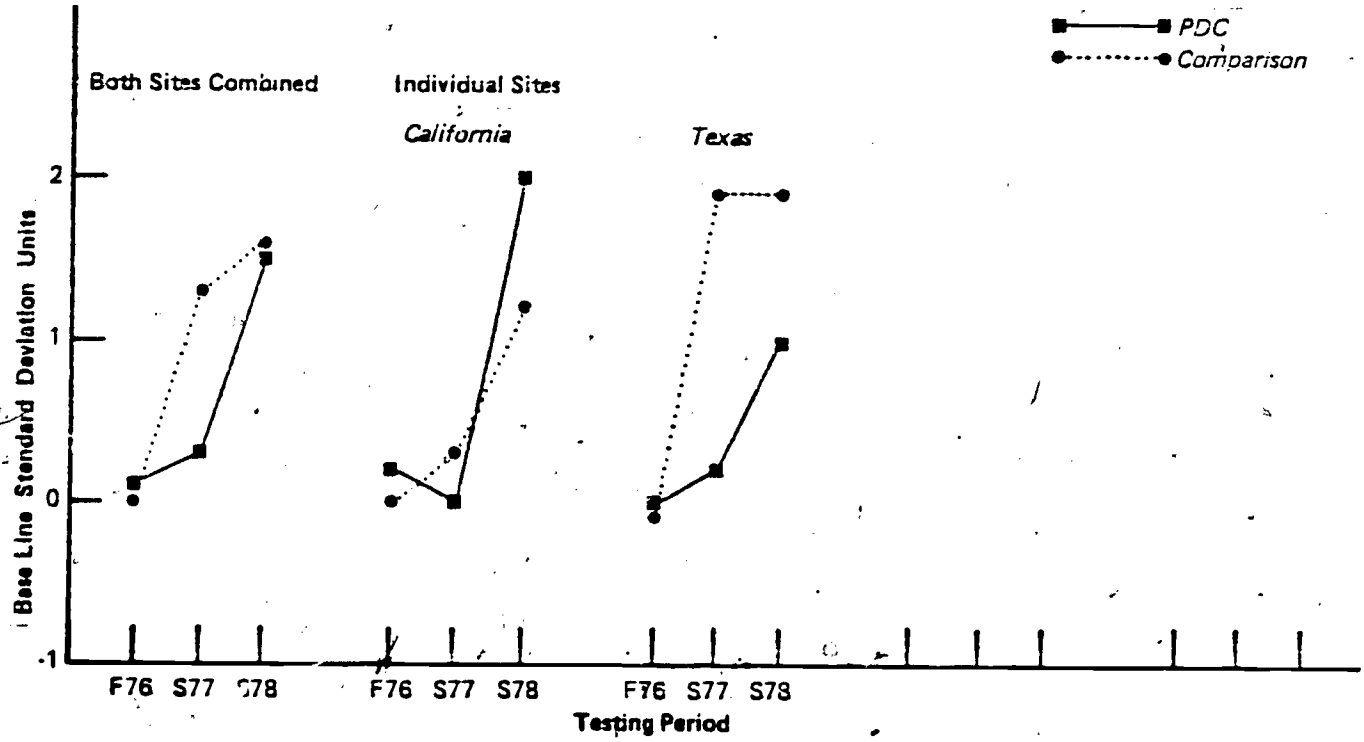


Figure 4b

Illustration of Main Program Effects and Program-by-Site Interactions for Preschool Interpersonal Problem Solving (PIPS) (Spanish-Dominant Analytic Sample)



Note: The points graphed represent group means adjusted for fall 1976 WPPSI score and for the effects of sex, ethnicity, and prior preschool experience. The means are standardized on the basis of the overall mean and standard deviation for the earliest testing period shown.

Figure 4c

Illustration of Main Program Effects and Program-by-Site Interactions for POCL-1: "Task Orientation" (Spanish-Dominant Sample)

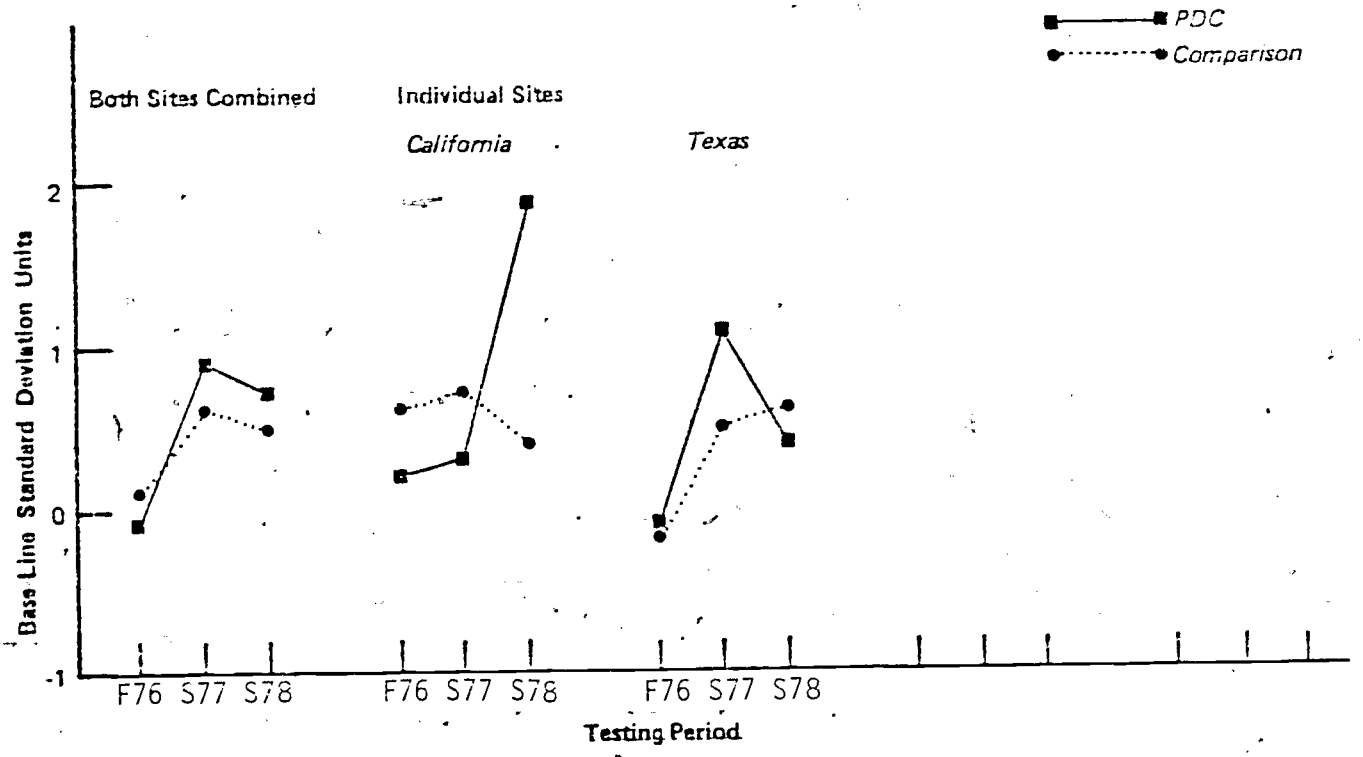
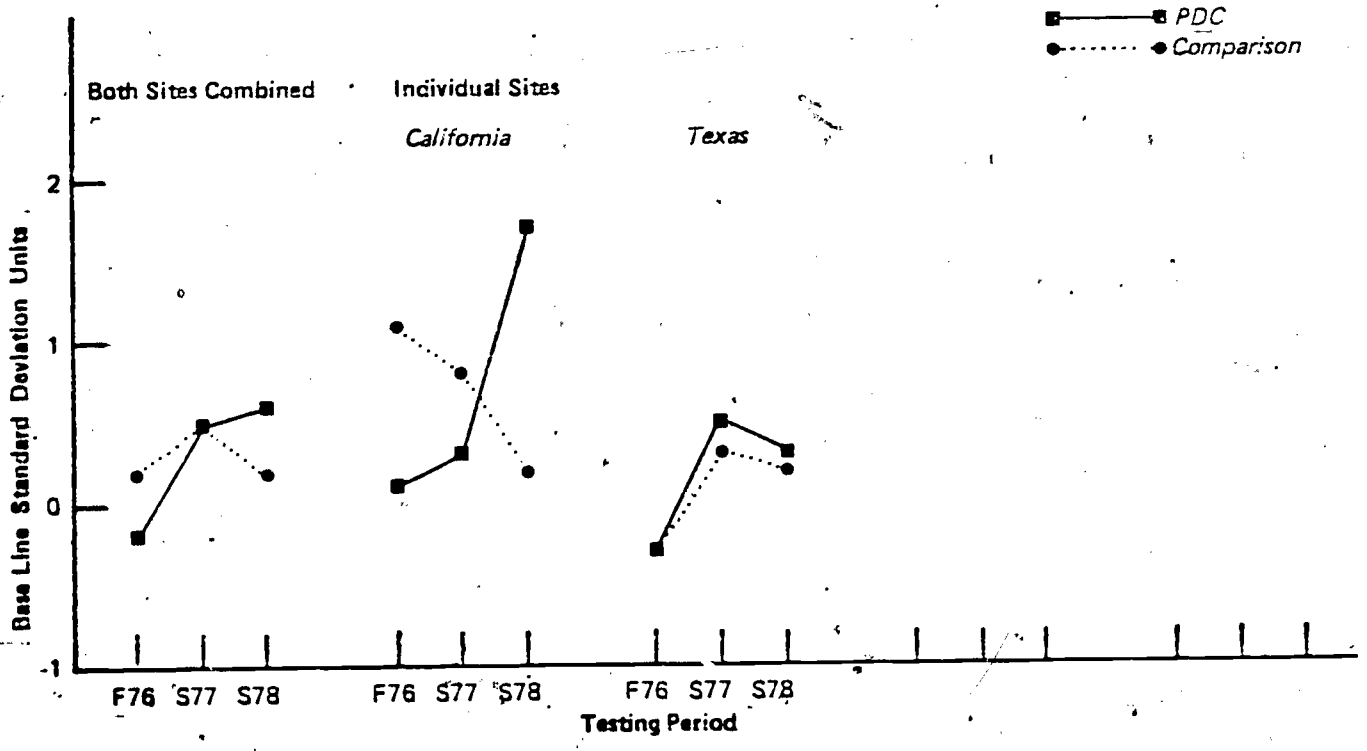


Figure 4d

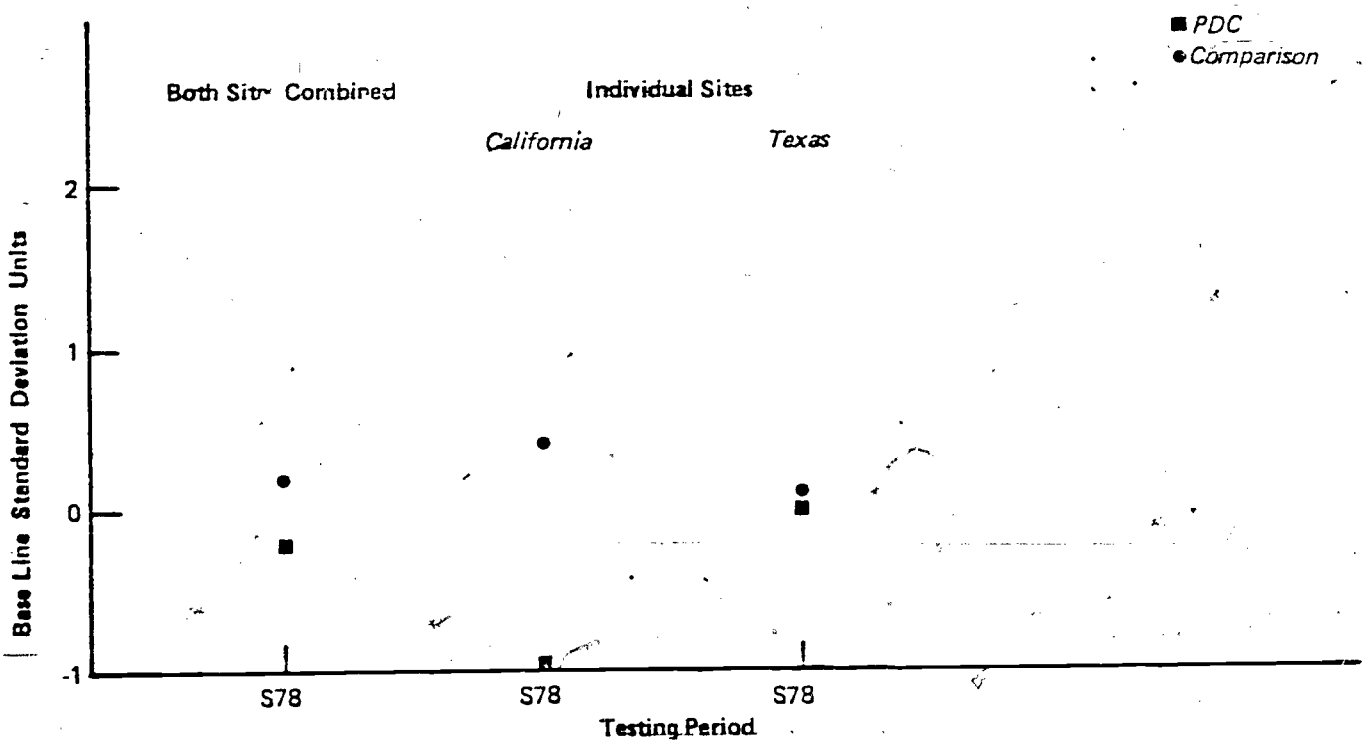
Illustration of Main Program Effects and Program-by-Site Interactions for POCL-2: "Sociability" (Spanish-Dominant)



Note: The points graphed represent group means adjusted for fall 1976 WPPSI score and for the effects of sex, ethnicity, and prior preschool experience. The means are standardized on the basis of the overall mean and standard deviation for the earliest testing period shown.

Figure 4e

Illustration of Main Program Effects and Program-by-Site Interactions for CRS-3: "Dependence" (Spanish-Dominant Analytic Sample)



Note: The points graphed represent group means adjusted for fall 1976 WPPSI score and for the effects of sex, ethnicity, and prior preschool experience. The means are standardized on the basis of the overall mean and standard deviation for the earliest testing period shown.

V

SUMMARY

Characteristics of the Sample

Comparability and representativeness of groups. The samples of English-dominant PDC and comparison children remain remarkably alike in terms of their characteristics at the time of program entry (fall 1976); attrition does not appear to have altered their comparability. Neither has it affected the basic makeup of either group with respect to the original PDC and comparison groups; the present samples of English-dominant children remain quite representative of the samples as they existed at the beginning of the Head Start year. The Spanish-dominant samples of PDC and comparison children also seem to satisfy the criteria of comparability and representativeness. But these groups are so small now that it is difficult to be confident of such statistical conclusions.

Adequacy of sample size. Among PDC children, the rate of attrition is higher than projected, but this is balanced by a rate among comparison children that is lower than projected. Thus, overall attrition at this point is quite close to our estimates. (The actual loss rate amounts to 36%, as against our March 1977 prediction of 39%.) However, since West Virginia will no longer be participating in PDC, this entire site will be lost to the evaluation in future years. Still, the number of PDC children likely to remain available through the end of third grade--about 20 per site--is large enough to represent the educational effects of PDC when sites are aggregated. The number of comparison children is likely to remain approximately equal to the number of PDC children, providing an adequate sample for statistical comparisons, at least, for analyses that focus on the English-dominant children. The Spanish-dominant sample has diminished to the point where conclusions based upon statistical analyses must acknowledge a large element of doubt.

Characteristics of the Measures

The basic battery. The instruments in the battery administered to English-dominant children continue to reflect high internal consistency and, altogether, a factor structure congruent with the intended structure. In the battery administered to Spanish-dominant children, some measures show signs of diminishing reliability. This may be a consequence of the diminishing size of this sample of children.

The Classroom Observation System, Phase I version. Our analyses of the data produced by this version of the observation system cast doubt on the system's suitability for measuring either individual behavior or classroom processes. Our speculative conclusion is that the brief samplings of behavior provided by this version of the system were not adequate to represent the complex nature of classroom processes. Thus, further interpretation of these data is unwarranted. A new classroom observation system has been developed for Phase II of the PDC evaluation, and this instrument is expected to provide deeper and more reliable insight into classroom processes than was possible with its predecessor.

Impact of the PDC Program

Impact at all sites considered in aggregation. Three series of analyses were undertaken to address three variations of the question, "Has PDC had an impact on children?" Even though the three analyses addressed somewhat different issues, and therefore would not necessarily produce identical results for any given measure of impact, it is when the analyses do agree that we can be most confident about our conclusions. For the sample of English-dominant children, the main analytic sample, there was no measure among all of those examined that seemed to be consistently affected by PDC (or by the treatments that comparison children experienced, either). The measure for which the strongest case can be made in regard to program impact is the "Aggressiveness" subscale of the Child Rating Scale. One analysis indicated that there was a general difference between PDC and comparison children on this variable, measured in spring 1977 and spring 1978; the other two analyses indicated that there was a group difference on this variable in at least some sites, but that the difference was not constant for all sites. It is difficult to say whether this result constitutes a positive impact of PDC, since it is uncertain how the quality of aggressiveness should be valued in the context of elementary education. On the basis of the analyses, it seems safe to say that PDC children rate lower in "Aggressiveness" than comparison children in some sites, if not in all sites. But is this result to be interpreted as positive, negative, or neither? Further, whatever the value placed on aggressiveness, one must ask whether the group difference found on this variable can plausibly be attributed to the impact of PDC. In the context of the finding that the PDC and comparison groups do not generally differ in a consistent way on the other variables examined, it seems quite possible that the difference between groups on "Aggressiveness" is due simply to chance rather than to a systematic program effect.

Impact at individual sites. While there is no persuasive evidence that the PDC program in general has had an educational impact on children as of their kindergarten year, it is possible that the program has begun to produce an impact in some sites. In Utah and West Virginia, particularly, we found differences favoring the PDC group on a number of measures. And there are other sites, in contrast, where the differences seem to favor the comparison children. We cannot state that these differences are statistically significant, however; nor can we say that they are necessarily due to the educational program as distinct from other factors. Our analyses did not pursue these issues to a high degree of certainty because of the lack of systematic information on the multitude of factors potentially influencing the impact PDC has. For this reason, the current report is intended only as an interim document for circulation within ACYF. In the future, the question of PDC's impact at individual sites will be investigated in ways intended to provide firmer statistical support for conclusions concerning site-to-site differences in program impact. And in documenting program impact, whether for all sites or for individual sites, we will also strive to determine just what dimensions of the program (including such factors as classroom environment, teacher attitudes and parent involvement) account for its impact.

APPENDIX A

Description of the Measures in the Spring 1978 Battery

	Order of Administration[1]
• Classroom-Level Measures	
PDC Classroom Observation System	1
• Social-Emotional Measures	
Preschool Interpersonal Problem-Solving Test (PIPS)	6
Pupil Observation Checklist (POCL)	9
PDC Child Rating Scale	-
• Psychomotor Measures	
Arm Coordination [McCarthy Scales of Children's Ability (MSCA)]	7
• Cognitive and Language Measures	
Bilingual Syntax Measure (BSM)	2
Block Design (WPPSI)[2]	administered fall 1976
Verbal Memory (MSCA)	3
Verbal Fluency (MSCA)	5
Draw-A-Child (MSCA)	4
Peabody Individual Achievement Test (PIAT)-Math & Reading Recognition Subtests	8

Each of these measures is described briefly below. For a more extensive review, see Interim Report II, Part B: Recommendations for Measuring Program Impact (1975).

---

[1]As noted in the text, the battery was administered in one or sometimes two sessions.

[2]WPPSI serves as an index of general cognitive aptitude, and was administered as a baseline measure at all sites in fall 1976. It was administered at only four sites in spring 1977. This was for the purpose of estimating its test-retest stability and sensitivity to change.

Bilingual Syntax Measure (Burt, Dulay and Hernandez Ch., 1975)[1]. This test is designed to measure children's oral proficiency in English and/or Spanish grammatical structures. Simple questions are used with cartoon-type colored pictures to provide a conversational setting for eliciting natural speech. An analysis of the child's response yields a numerical indicator and a qualitative description of the child's structural language proficiency in standard English or standard Spanish. Responses are written down verbatim.

Children who show facility in both Spanish and English receive both versions of the BSM. The order in which the two versions are administered is controlled so that during any single testing period half the children receive the Spanish version first and half receive the English version first.

McCarthy Scales of Children's Abilities (McCarthy, 1972)[2]. These subtests consist of a series of tasks tapping problem-solving, psychomotor, and conceptual abilities, and are similar to the Wechsler Scales, but with emphasis on age-related maturational indicators.

- Verbal Memory. The child is asked to repeat sequences of words (Verbal Memory-1) and to repeat or retell as much as possible of a one-paragraph story (Verbal Memory-3). McCarthy's Scale 2 is not used in the PDC evaluation.

- Verbal Fluency. The child is asked to name as many members of specific categories (e.g., animals) as he/she can.

- Arm Coordination. Child bounces a rubber ball, catches a beanbag, and throws a beanbag through a hole in a target.

- Draw-A-Child. Child draws a picture of a child of the same sex.

---

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

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

PDC Child Rating Scale (High/Scope Foundation, unpublished). This instrument, designed as a measure of the construct "social competence," is completed by each child's classroom teacher. The scale consists of items that refer to aspects of school behavior, such as "Uses words or wits to influence others," and the teacher rates the frequency with which the child demonstrates this behavior ("V y frequently" to "Rarely"). As noted in the text, the 39 items of the original Child Rating Scale have been reduced to 17 (comprising the scales "Self-Assurance," "Aggressiveness," and "Dependence") and three new items have been added to constitute an "Academic Motivation" scale (to be administered from spring 1979 onward).

PDC Classroom Observation System (High/Scope Foundation, unpublished). The PDC observation system was developed to provide information about children's classroom behavior along dimensions pertinent to the social-emotional goals of Project Developmental Continuity. The system focuses on aspects of an individual child's behavior, verbal or nonverbal, that reflect the child's attitude toward himself, and on the child's social competence as demonstrated in his interaction with peers and adults.

Using a time-sampling method, trained observers observe each child for five minutes at two different times during the day and code the observed behavior into four general categories: "noninvolved," "involved," "interacts with peer," and "interacts with adult." A fifth category, "activity level," is included to provide information concerning the context in which these behaviors were observed. Each of these categories includes subcategories that are designed to identify the frequency and nature of specific behaviors within the general category.

Peabody Individual Achievement Test (Dunn and Markwardt, 1970)[1]. The math and reading recognition subtests of this individually administered instrument were used to assess these two dimensions of school achievement. The math subtest taps skills such as computation, mastery of ordination, shape discrimination, and understanding of key terms, such as "double" and "youngest". The examiner reads each math question aloud and the child responds by pointing to one of four possible answers displayed on a plate in the administrators's notebook. The

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

reading recognition subtest consists of letter recognition tasks, which involve the matching and naming of alphabetic letters, and word recognition tasks, which consist of reading written words aloud.

Preschool Interpersonal Problem-Solving Test (Shure and Spivack, 1974)[1]. The PIPS attempts to assess the child's ability to name alternative solutions to a life-related problem--that of obtaining a toy from another child. Paper cut-outs of boys, girls and toys are used in presenting the problem. Among inner city four-year-olds attending the Philadelphia Get Set day care program, those judged as better-adjusted by their teachers were able to conceptualize a greater number and a wider range of alternative solutions to real-life problems than were their more poorly adjusted classmates.

Pupil Observation Checklist (High/Scope Foundation, unpublished). This is a rating scale consisting of twelve 7-point bipolar adjectives derived from a similar scale used in the Home Start evaluation.[2] The tester rates each child using this instrument after he or she has administered all the other measures in the battery to the child. The two subscales of the POCL, "Task Orientation" (9 items) and "Sociability" (3 items) were derived by factor analysis. The factor structure has been replicated in a number of independent studies.

Wechsler Preschool and Primary Scale of Intelligence, Block Design Subtest (Wechsler, 1967)[3]. The task requires reproducing (constructing) designs with flat colored blocks, either from the examiner's model or from a picture on a card. The measure taps problem-solving abilities, flexibility of response style, visual-motor organization, and execution. The WPPSI was administered only in fall 1976, with the intention that this measure would continue to serve as a covariate in future analyses.

---

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

[2]Love, J., et al. National Home Start Evaluation Interim Report VII. Ypsilanti, MI: High/Scope Foundation, March 1976.

[3]Wechsler, D. Wechsler Preschool and Primary Scale of Intelligence: Manual. New York: Psychological Corporation, 1967.

APPENDIX B

Significance of the Contributions of Covariates and Design Factors in Impact Analysis Series 1, 2, and 3.

For the sake of documenting the roles of the covariates and design factors entering into the analyses of PDC's impact, Tables B-1, B-2 and B-3 indicate the significance levels attained by each term in the model for each analysis. It would be preferable to include in these tables estimates of the magnitude of the contribution (e.g., R^2 coefficients), but these estimates are unfortunately not recoverable from the output of the computer program used.

Included additionally in Table B-3 is the significance level associated with a test for linear trend in each of the analyses performed. This is a test of the hypothesis that the dependent measures describe a line that increases or decreases (rather than remaining constant) over time.

Table B-1

Significance of the Contributions of Covariates and Design Factors in Analysis Series #1
(English-Dominant Analytic Sample)

	BSM-English	Verbal Fluency	Verbal Memory-1	Verbal Memory-3	Arm Coordination	Draw-a-Child	PIPS	POCL-1: "Task Orientation"	POCL-2: "Sociability"	PIAT-Math	PIAT-Reading	CRS-1: "Self-Assurance"	CRS-2: "Aggressiveness"	CRS-3: "Dependence"
Covariates														
WPPSI (F76)	*	*	**	NS	NS	**	NS	*	NS	***	***	***	NS	*
F76 score on the dependent variable	***	***	***	***	***	***	***	***	***	a	a	a	a	a
Design Factors														
Program	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Site	***	***	*	NS	NS	NS	*	***	***	**	***	**	NS	***
Ethnicity	***	***	NS	NS	NS	NS	NS	*	**	**	*	NS	*	NS
Sex	NS	NS	NS	NS	***	*	NS	NS	NS	NS	*	***	**	NS
Prior preschool experience	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Program-by-site interaction	NS	NS	NS	**	NS	NS	NS	*	NS	**	***	NS	**	NS

[a]Measure not administered at this timepoint. WPPSI alone was used as a covariate in this analysis.

NS = nonsignificant

*probability of associated F ratio < .05

**probability of associated F ratio < .01

***probability of associated F ratio < .001

Table B-2

Significance of the Contributions of Covariates and Design Factors in Analysis Series #2
(English-Dominant Analytic Sample)

	BSM-English	Verbal Fluency	Verbal Memory-1	Verbal Memory-3	Arm Coordination	Draw-a-Child	PIPS	POCL-1: "Task Orientation"	POCL-2: "Sociability"	PIAT-Math	PIAT-Reading	CRS-1: "Self-Assurance"	CRS-2: "Aggressiveness"	CRS-3: "Dependence"
Covariates														
WPPSI (F76)	NS	NS	*	NS	NS	**	NS	NS	NS			*	NS	NS
F76 score on the dependent variable	**	***	***	**	***	***	*	NS	**			a	a	a
S77 score on the dependent variable	***	***	***	***	***	***	***	***	***			***	***	NS
Design Factors														
Program	NS	NS	NS	NS	NS	NS	NS	NS	NS			*	NS	NS
Site	**	*	NS	NS	NS	NS	**	***	***			**	NS	***
Ethnicity	***	***	NS	NS	NS	NS	NS	NS	*			NS	NS	NS
Sex	NS	*	NS	NS	***	NS	NS	NS	NS			**	*	NS
Prior preschool experience	NS	NS	NS	NS	NS	NS	NS	NS	NS			NS	NS	NS
Program-by-site interaction	NS	NS	NS	**	NS	NS	NS	NS	NS			NS	***	NS

----- Analysis not applicable -----

----- Analysis not applicable -----

[a]Measure not administered at this timepoint. WPPSI and S77 score on the same CRS scale were used as the covariates in this analysis.

NS = nonsignificant

*probability of associated F ratio < .05

**probability of associated F ratio < .01

*probability of associated F ratio < .001

Table B-3

Significance of the Contributions of Covariates and Design Factors in Analysis Series #3
(English-Dominant Analytic Sample)

	BSM-English	Verbal Fluency	Verbal Memory-1	Verbal Memory-3	Arm Coordination	Draw-a-Child	PIPS	POCL-1: "Task Orientation"	POCL-2: "Sociability"	PIAT-Math	PIAT-Reading	CRS-1: "Self-Assurance"	CRS-2: "Aggressiveness"	CRS-3: "Dependence"
Covariates														
WPPSI (F76)	***	***	***	***	***	***	**	***	*	----- Analysis not applicable -----	----- Analysis not applicable -----	***	NS	*
Design Factors														
Program	NS	NS	NS	NS	NS	NS	NS	NS	NS			NS	*	NS
Site	***	***	*	**	**	***	***	***	NS			**	NS	*
Ethnicity	***	NS	*	NS	**	NS	NS	NS	NS			NS	NS	NS
Sex	NS	NS	**	NS	***	***	**	**	*			**	*	NS
Prior preschool experience	NS	NS	NS	NS	NS	NS	NS	NS	NS			NS	*	NS
Program-by-site interaction	NS	NS	*	NS	NS	NS	NS	NS	NS			NS	NS	NS
Growth Curve														
Significant linear time trend indicated?	YES	YES	YES	YES	YES	YES	YES	YES	NO			NO	YES	YES

NS = nonsignificant

*probability of associated F ratio < .05

**probability of associated F ratio < .01

***probability of associated F ratio < .001